

```
1 CGAGTGCCT GCTGAGTCCT GTAGATAAAG CCGCCAACCC CGGGGACTGG
51 TGTCTCCTGA GTGACCGTGC AGCCGTGGGC GCCATAGAAA GCAGAGAAGG
101 CAGTGAACCTT CGACCACTTC CAGATCCTTC GGGCCATTGG GAAGGGCAGC
151 TTTGGCAAGG TGTGCATTGT GCAGAAGCGG GACACGGAGA AGATGTACGC
201 CATGAAGTAC ATGAACAAGC AGCAGTGCAT CGAGCGCGAC GAGGTCCGCA
251 ACGTCTTCCG GGAGCTGGAG ATCCTGCAGG AGATCGAGCA CGTCTTCTG
301 GTGAACCTCT GGTACTCCTT CCAGGACGAG GAGGACATGT TCATGGTCTG
351 GGACCTGCTA CTGGCGGGG ACCTGCGCTA CCACCTGCAG CAGAACGTGC
401 AGTTCTCCGA GGACACGGTG AGGCTGTACA TCTGCGAGAT GGCCTGGCT
451 CTGGACTACC TGCGCGGCCA GCACATCATC CACAGAGATG TCAAGCCTGA
501 CAACATTCTC CTGGATGAGA GAGGACATGC ACACCTGACC GACTTCAACA
551 TTGCCACCAT CATCAAGGAC GGGGAGCGGG CGACGGCATT AGCAGGCACC
601 AAGCCGTACA TGGCTCCGGA GATCTTCCAC TCTTTTGTCA ACGGCGGGAC
651 CGGCTACTCC TTCGAGGTGG ACTGGTGGTC GGTGGGGGTG ATGGCCTATG
701 AGCTGCTGCG AGGATGGAGG CCCTATGACA TCCACTCCAG CAACGCCGTG
751 GAGTCCCTGG TGCAGCTGTT CAGCACCGTG AGCGTCCAGT ATGTCCCCAC
801 GTGGTCCCAAG GAGATGGTGG CTTTGTGCG GAAGCTCCTC ACTGTGAACC
851 CCGAGCACCG GCTCTCCAGC CTCCAGGACG TGCAGGCAGC CCGGCGCTG
901 CCGGCGGTGC TGTGGGACCA CCTGAGCGAG AAGAGGGTGG AGCCGGGCTT
951 CGTGCCCAAC AAAGGCCGTC TGCACTGCGA CCCCACCTTT GAGCTGGAGG
1001 AGATGATCCT GGAGTCCAGG CCCCTGCACA AGAAGAAGAA GCGCTGGCC
1051 AAGAACAAAGT CCGGGGACAA CAGCAGGGAC AGCTCCAGT CCGAGAATGA
1101 CTATCTTCAA GACTGCCTCG ATGCCATCCA GCAAGACTTC GTGATTTTGA
1151 ACAGAGAAAA GCTGAAGAGG AGCCAGGACC TCCCGAGGGA GCCTCTCCCC
1201 GCCCTGAGT CCAGGGATGC TGCGGAGCCT GTGGAGGACG AGGCGGAACG
1251 CTCCGCGCTG CCCATGTGCG GCCCATTTTG CCCCTCGGCC GGGAGCGGCT
1301 AGGCCGGGAC GCCCGTGGTC CTCACCCCTT GAGCTGCTTT GGAGACTCGG
1351 CTGCCAGAGG GAGGGCCATG GGGCGAGGCC TGGCATTAC GTTCCCACCC
1401 AGCCTGGCTG GCGGTGCCCA CAGTGCCCGG GACACATTTC ACACCTCAGG
1451 CTCGTGGTGG TGCAGGGGAC AAGAGGCTGT GGGTGCAGGG GACACCTGTG
1501 GAGGGCATTT CCCGTGGGCC CCCGAGACCC GCCTAGATGG AGGAAGCGCT
1551 GCTGGGCGCC CTCTTACCGC TCACGGGGAG CTGGGGCCAT GGATGGGACA
1601 GGAGTCTTTG TCCCTGCTCA GCCCGGAGGC TGTGCACGGC CCTCGTCACA
1651 AGGTGACCCT TGCAGCACAG GCCGCGGGTG CCCGAGGCTC GGCTCAGTTC
1701 TTGGAGGTCA AGGGCATGGG TTGGGGTAGT GGGTGGGGAG GTGAATGTTT
1751 TCTAGAGATT CAAACTGCTC CAGCAATTTT TGTATAGTTT TCACCTCTGA
1801 GAATTACAAT GTGAGAACCG CACAAAAAAA AAAAAA AAAA
1851 AAAAAA AAAA
(SEQ ID NO: 1)
```

FEATURES:

5' UTR: 1 - 192
Start Codon: 193
Stop Codon: 1300
3' UTR: 1303

Homologous proteins:

Top 10 BLAST Hits:

Sequences producing significant alignments:	Score (bits)	E value
CRA 103000001515936 /altid=gi 10946600 /def=ref NP_067277.1 hy...	760	0.0
CRA 120000042903164 /altid=gi 13358640 /def=dbj BAB33045.1 (AB...	733	0.0
CRA 870000000001314 /altid=gi 8923754 /def=ref NP_060871.1 gene...	556	e-157
CRA 870000000001426 /altid=gi 7161864 /def=emb CAB76566.1 (AJ25...	554	e-156
CRA 108000024647823 /altid=gi 12730486 /def=ref XP_003392.2 ge...	423	e-117
CRA 18000005184360 /altid=gi 7505957 /def=pir T23688 hypotheti...	335	1e-90
CRA 18000005004115 /altid=gi 1730069 /def=sp P54644 KRAC_DICDI ...	217	4e-55
CRA 18000004912236 /altid=gi 464395 /def=sp P28178 PK2_DICDI PR...	203	8e-51
CRA 18000004910302 /altid=gi 1170689 /def=sp P42818 KPK1_ARATH ...	202	1e-50
CRA 89000000197925 /altid=gi 7295638 /def=gb AAF50945.1 (AE003...	201	2e-50

FIGURE 1A

Docket No.: CL001078-DIV
Serial No.: TO BE ASSIGNED
Inventors: Ellen BEASLEY et al.
Title: ISOLATED HUMAN KINASE ...

EST:

Sequences producing significant alignments:

	Score (bits)	E value
gi 13032240 /dataset=dbest /taxon=960...	1348	0.0
gi 6588496 /dataset=dbest /taxon=9606 ...	1021	0.0
gi 883123 /dataset=dbest /taxon=9606 /...	702	0.0
gi 946492 /dataset=dbest /taxon=9606 /...	236	1e-59

EXPRESSION INFORMATION FOR MODULATORY USE:

gi|13032240 prostate
gi|6588496 /lung
gi|883123 /whole brain
gi|946492 / Adult brain

Tissue Expression

whole brain

FIGURE 1B

Docket No.: CL001078-DIV
Serial No.: TO BE ASSIGNED
Inventors: Ellen BEASLEY et al.
Title: ISOLATED HUMAN KINASE ...

1 MYAMKYMKNQ QCIERDEVNR VFRELEILQE IEHVFLVNLW YSFQDEEDMF
51 MVVDLLLGDD LRYHLQQNVQ FSEDTVRLYI CEMALALDYL RGQHIIHRDV
101 KPDNILLDER GHAILTDFNI ATIIKDGERA TALAGTKPYM APEIFHSFVN
151 GGTGYSFEVD WWSVGVMAYE LLRGWRPYDI HSSNAVESLV QLFSTVSVQY
201 VPTWSKEMVA LLRKLLTVNP EHRLSSLQDV QAAPALAGVL WDHLSEKRV
251 PGFVPNKGRL HCDPTFELEE MILESRLPHK KKKRLAKNKS RDNSRDSSQS
301 ENDYLQDCLD AIQQDFVIFN REKLKRSQDL PREPLPAPES RDAAEPVEDE
351 AERSALPMCG PICPSAGSG
(SEQ ID NO: 2)

FEATURES:

Functional domains and key regions:

[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
N-glycosylation site

288-291 NKSR

[2] PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site

Number of matches: 2

1 75-77 TVR
2 245-247 SEK

[3] PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site

Number of matches: 4

1 42-45 SFQD
2 226-229 SLQD
3 298-301 SQSE
4 300-303 SEND

[4] PDOC00008 PS00008 MYRISTYL
N-myristoylation site

152-157 GTGYSF

[5] PDOC00100 PS00108 PROTEIN_KINASE_ST
Serine/Threonine protein kinases active-site signature

95-107 IIHRDVKPDNILL

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	185	205	0.727	Putative

FIGURE 2A

BLAST Alignment to Top Hit:

>CRA|103000001515936 /altid=gi|10946600 /def=ref|NP_067277.1|
hypothetical serine/threonine protein kinase [Mus
musculus] /org=Mus musculus /taxon=10090 /dataset=nraa
/length=488
Length = 488

Score = 760 bits (1942), Expect = 0.0
Identities = 371/399 (92%), Positives = 381/399 (94%), Gaps = 1/399 (0%)
Frame = +1

Query: 103 VNFDFHQLRAIGKGSFGKVCIVQKRDEKMYAMKYMKNQCCIERDEVRNVFRELEILQE 282
Sbjct: 89 VNFDFHQLRAIGKGSFGKVCIVQKRDEKMYAMKYMKNQCCIERDEVRNVFRELEILQE 148
Query: 283 IEHVFLVNLWYSFQDEEDMFVVDLLGGDLRYHLQQNVQFSEDTVRLYICEMALALDYL 462
Sbjct: 149 IEHVFLVNLWYSFQDEEDMFVVDLLGGDLRYHLQQNVQFSEDTVRLYICEMALALDYL 208
Query: 463 RGQHIHRDVKPDNILLDERGHAHLTDFNIATIIKDGERATALAGTKPYMAPEIFHSFVN 642
Sbjct: 209 RSQHIHRDVKPDNILLDE+GHAHLTDFNIATIIKDGERATALAGTKPYMAPEIFHSFVN 268
Query: 643 GGTGYSFEVDWWSVGVMAYELLRGWRPYDIHSSNAVESLVQLFSTVSVQYVPTWSKEMVA 822
Sbjct: 269 GGTGYSFEVDWWSVGVMAYELLRGWRPYDIHSSNAVESLVQLFSTVSVQYVPTWSKEMVA 328
Query: 823 LLRKLTLVNPEHRLSSLQDVQAAPALAGVLWDHLSEKRVEPGFVPNKGRLHCDPTFELEE 1002
Sbjct: 329 LLRKLTLVNPEHRLSSLQD+Q AP+LA VLWD LSEK+VEPGFVPNKGRLHCDPTFELEE 388
Query: 1003 MILESRLPHKKKKRLAKNKS RDNSRDSSQSENDYLQDCLDAIQQDFVIFNREKLKRSQDL 1182
Sbjct: 389 MILESRLPHKKKKRLAKNKS RD+SRDSSQSENDYLQDCLDAIQQDFVIFNREKLKRSQ+L 448
Query: 1183 PREPLPAPESRDAAEPVED-EAERSALPMCGPICPSAGS 1296
EP P PE+ D + D EAE +ALPMCG ICPS+GS
Sbjct: 449 MSEPPPGPETSMDTSTADSEAEPTALPMCGSICPSSGS 487 (SEQ ID NO: 4)

>CRA|120000042903164 /altid=gi|13358640 /def=dbj|BAB33045.1|
(AB056389) hypothetical protein [Macaca fascicularis]
/org=Macaca fascicularis /taxon=9541 /dataset=nraa
/length=368
Length = 368

Score = 733 bits (1872), Expect = 0.0
Identities = 358/369 (97%), Positives = 361/369 (97%)
Frame = +1

Query: 193 MYAMKYMKNQCCIERDEVRNVFRELEILQEIEHVFLVNLWYSFQDEEDMFVVDLLGGD 372
Sbjct: 1 MYAMKYMKNQCCIERDEVRNVFREL ILQEIEHVFLVNLWYSFQDEEDMFVVDLLGGD 60
Query: 373 LRYHLQQNVQFSEDTVRLYICEMALALDYL RGQHIHRDVKPDNILLDERGHAHLTDFNI 552
Sbjct: 61 LRYHLQQNVQFSEDTVRLYICEMALALDYL CGQHIHRDVKPDNILLDERGHAHLTDFNI 120
Query: 553 ATIIKDGERATALAGTKPYMAPEIFHSFVNGGTGYSFEVDWWSVGVMAYELLRGWRPYDI 732
Sbjct: 121 ATIIKDGERATALAGTKPYMAPEIFHSFVNGGTGYSFEVDWWS+GVMAYELLRGWRPYDI 180
Query: 733 HSSNAVESLVQLFSTVSVQYVPTWSKEMVALLRKLTLVNPEHRLSSLQDVQAAPALAGVL 912
Sbjct: 181 HSSNAVESLVQLFSTVSVQYVPTWSREMVALLRKLTLVNPEHRLSSLQDVQAAPALAGVL 240
Query: 913 WDHLSEKRVEPGFVPNKGRLHCDPTFELEEMILESRLPHKKKKRLAKNKS RDNSRDSSQS 1092
Sbjct: 241 WGHLEKRVEPDFVPNKGRLHCDPTFELEEMILESRLPHKKKKRLAKNKS RDNSRDSSQS 300

FIGURE 2B

Docket No.: CL001078-DIV
Serial No.: TO BE ASSIGNED
Inventors: Ellen BEASLEY et al.
Title: ISOLATED HUMAN KINASE ...

Query: 1093 ENDYLQDCLDAIQQDFVIFNREKLKRSQDLPREPLPAPESRDAAEPVEDEAERSALPMCG 1272
ENDYLQDCLDAIQQDFVIFNREKLKRSQDL P EPLPAPE RDAAEPVEDE E+SALPMCG
Sbjct: 301 ENDYLQDCLDAIQQDFVIFNREKLKRSQDLPSEPLPAPEPRDAAEPVEDE-EQSALPMCG 359

Query: 1273 PICPSAGSG 1299
PICPSAGSG
Sbjct: 360 PICPSAGSG 368 (SEQ ID NO: 5)

Hmmer search results (Pfam):

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
PF00069	Eukaryotic protein kinase domain	197.7	1.8e-55	1
CE00359	E00359 bone_morphogenetic_protein_receptor	9.8	0.044	1
CE00022	CE00022 MAGUK_subfamily_d	9.2	0.013	1
CE00031	CE00031 VEGFR	1.6	1.3	1
CE00292	CE00292 PTK_membrane_span	-92.7	0.0016	1
CE00291	CE00291 PTK_fgf_receptor	-116.1	0.041	1
CE00287	CE00287 PTK_Eph_orphan_receptor	-125.0	2.8	1
CE00286	E00286 PTK_EGF_receptor	-131.8	0.0056	1
CE00290	CE00290 PTK_Trk_family	-206.7	0.58	1
CE00016	CE00016 GSK_glycogen_synthase_kinase	-248.2	0.0061	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00022	1/1	86	107 ..	133	154 ..	9.2	0.013
CE00359	1/1	95	144 ..	272	326 ..	9.8	0.044
CE00031	1/1	79	171 ..	1051	1141 ..	1.6	1.3
CE00286	1/1	3	212 ..	1	263 []	-131.8	0.0056
CE00287	1/1	7	212 ..	1	260 []	-125.0	2.8
CE00290	1/1	4	215 ..	1	282 []	-206.7	0.58
PF00069	1/1	2	228 ..	26	271 ..	197.7	1.8e-55
CE00292	1/1	5	233 ..	1	288 []	-92.7	0.0016
CE00291	1/1	1	233 []	1	285 []	-116.1	0.041
CE00016	1/1	1	303 []	1	433 []	-248.2	0.0061

FIGURE 2C

```

1 GATGGAGCCC AGGCCACATG CTTGGGAAAG CGGGTGGTCC TGGGTGAGCC
51 TGTCTGGCAG GAGGGTGAGG TCCTGTCTGG AGTAGACACC CGGTCTGTGT
101 CCGCCGCACA TGTGACAGGG CCCAGAGGGC ACGGAAGACC CAGGTGCCCT
151 GTCCCTGGT GAGTGGGCTG TGGGGGTTGA GCACCCCAAG GGAAGGCGTG
201 TTCTTCCAGA GATTCCTCCG TAAGAGCTGA GCTGCATCGT GAGCAGGAGG
251 GGCAGAGGGA GCCTGGAGAG GGTGGGTCG CCTGGGCAGG CAGCTTTCTG
301 GGGTGGTGCA CAAGCGACAG AAAGCCTGAG GCTTGGGACC GGCTGAAGAC
351 GGCCTTGGGC TGCGGGTGCT GCGCGCGCTC CAGGCAGAGG TGGGCTCTGC
401 AGCCCCCCCC CCCCCGCCAG CTTCCCCCCC AGCAGCAAGA GCCCTGGCCC
451 CAGCCAGTGG CCCCAGGCA GGCTCCTGAG GGGCAGAGGG TGAGGGCCGG
501 ATTTTCCACC ATATTTGTCT CACAGCCTGT CTGGTCCAG CCCAGGGCA
551 ACAAACAGCC TTTCTGGAGC AGTTTCCAGA CCTGCAGTGG CCGCTTGAG
601 CCTGCAGTGA CCGTCTGCAG GAGGCCGCGG GTGCTGGGGC TGGCGCAGGA
651 AAGCACCGTT GCTTCTGCGC CTGTGCAGAG TGAGGCTGGG GCTTCCATCC
701 CGGGCACGGG ACTCCTCGGC CTCCTGCGGC CGTGTGCATG GGAGGAAGGC
751 CGTGCTGCCG AGCCACGCGA CTCTGCCCCG TTGGCAGTGG GAAGCGGCAG
801 GAGGGGGTCC TGCCAGGGGC AGCCAGGGGC TGCTGCAGT TACGCTACT
851 GTCATTCTGA AACCCTCAAC TGGCTTTCAA AATAACAATT TAAAAAATGG
901 TCATAGGAAA TGCAGGAAGT TCAGGAGAAA TCCCGCCCCG CCCCGCCCTC
951 CCAAGGCGGC CTCGTTCAAG CTTAGTCTCC GTCTGTTCTC GGGCTGCCTG
1001 CAGCTGCCCC CGCTCCTCAG CAGGTGTGGC CGCTGTTCAG GGAGCCCCCA
1051 TCAGCACACG CGCTCTGGG CAGCCCCCAA CACAAGGCTC TTCTGTCTCC
1101 TTCAGGCTCA GCTTCCCCCT CCCACCGGGC AGGGAGGTGC TGAGGCCACG
1151 CCTGTTGTCA GCTTCTGGA GAGGCCATAC TTAGAGCCA TGGTGCCAGG
1201 CCAAGCACT GTCTCCGCA CCACAGTGCT GCAGCCGGAT CCACCCAGGC
1251 CACCGCTTGG CACCATCAGA CATGCTTTCT TAGTTTGGG CAGCCCGGTG
1301 CTCTGTGCAC AGTGTGACTC CAGTGGCCC CTTGCGGGAG GGAGGGTCAC
1351 CGCTTTGCTT CACAAAAGGA CTGAGTCTCA GGGAGGGGTC TCCAGTAAGG
1401 GCCCCGGAGC CAGGATGTGA CTGAGACAGG TGCTCCAGG GCCACACACT
1451 AACCAGTACA GGAACCTCTG GGGGCAAGT CATGGCCTCC AAAACACCCA
1501 CATCCAAAT CCAGAAACAA CATGTTACCA TTCGTGGCAG AGAGGAATGA
1551 AGATTGCAGA TTGAATTAAG GTTGTTAATC ATCTGACTTA ATTTTITTAG
1601 AGACAGGGTC TCGCCCTGTC ACCCAGGCTG AGTTCAGTGG TACAGTCGTG
1651 GCTCACTGCA GCCTCAACCT CTCAGGCTCA AGCAATCCTC CCTTCTCAGT
1701 CTCTGATTGA GTCTGGGACTA CGGGTGTGTG CCACCACGCT TTCCGGGCAT
1751 GCAGCCAGGA GCCCAGGGCC ATCTGTGGCC CACCTTGAGA TCCAGAATCA
1801 TCCATTTCTT CCAGGCCCCC TGCTGGGCTC CAATCTCTTG AGGACCAGAG
1851 AGCAGAGGTT GTGGAAGGCC TTGGAACCGG GTCTGGATTA CCTGTCCTGG
1901 GAAGGTCTCT CCAACCTGA GTGTACAGCA GGGGTAGCT CTGCTGCTCA
1951 CAATTTTGTG CCTTAATTCC TGGCTTCCCT TTGGGGATCT TCATCCTCAA
2001 TTCTGATTGA CACCTTGGC CACAAGGGAC CCCCTGCTC ATTGATGCTT
2051 CTCACCCGTC ACCTCACTCT CATCCTCACT GCTAAGCAAT TAGCCGTGTG
2101 TTTGCGGCAT CAGTGTGAC ACCGATGATC CATGCTCAGA GGGTACAGGC
2151 CTGAAGAGCT ATGTGGGGAC TGGCGCCCCG GAGGGGGTCC CGCTGTGGTG
2201 GCAGCGGTGG CCCCCAAGCC CCACGCTCAC TCTGTGTGTC TCCTTGACAGG
2251 ACAGGGTGAG GGGCCTTGGC CTCACGGTGT TGAGACGGGA GTCGGTCTCT
2301 AGAGTGTGGA GTGATGTGCG TCCAGGGTGA AGCTGCAGCC ATGTTGATG
2351 AGGCTTCTCT AGCGCAGGGC TCTGTGCTCA GTGGGTTTTT GCATTCAATC
2401 CCCAGTACCC CTCGGGGCTG CTTGTGTGTC CAAGCCCTGG AAAAGGATGT
2451 TGGGGTTTAG GAAGGCAAAA CTCCATGCCC AGGTCGTGGC TGGTGAGGGG
2501 CGCTCCCCGC ACAACCAGCC ACTGCTTGGC TCCAACCTAC GCCTGGATGC
2551 TGTTAGGCTG GACCCTGTCT GTTTGCAGAT AGCGCTGTT GACAGATGTG
2601 TCCTGCTGCA GACTTGAAAC GCAGGACTGA GTCTCAGGGA GGGGTCTTCA
2651 CTAACAGCCA TGGAGCCAGG ATGTCACTGA GACAGGTGCC GCCAGGGCCA
2701 CACACTAACC AGTACAGGAA CCTCTGGGGG GCAGAATCAC GGCCTCCAAA
2751 ACACCCACAT CTTAATCCCC AGAGCAAATA TGTTACCATA TGTGGCAGAG
2801 AGGAATGCAC CTGCCGCGTC TAGGAGGCAG AGGGGCTGCG GCGCGTCCCC
2851 AGGTGTCCCC TTGTGCTCTG ACGATGCGTC CCCAGGCATC CCCATGCATT
2901 TCCAGGTGTC CCCACGTGTC CCCAGGCGTC CTCAGGCGTC CTGCCGAACA
2951 GCCCTGTGCC CTCCAGGTGT GCATTGTGCA GAAGCGGGAC ACGGAGAAGA
3001 TGTACGCCAT GAAGTACATG AACAAGCAGC AGTGATCGA GCGCGACGAG
3051 GTCCGCAACG TCTTCCGGGA GCTGGAGATC CTGCAGGAGA TCGAGCACGT
3101 CTTCTGGTG AACCTCTGGT GAGCCTGCCA TGGCCTCGCT GCAGAAAGAC
3151 TTACCGTCTT GAAGCCGGGA AGGCACTGGC TATCCTCTCC TGCCCTTGGT
3201 GTCTCTTGGC CACTGTGCTC CGGAGATGCC TCTGCCACCC ACAGGCCCTC
3251 TCTAGCCCTC CTCATCTATC CCCCTTCTCT TGCCCCAGGC CTGGCAGTGG
3301 CCCAGGTGGC CATGACATGC TGGGGTTGGT TAATGCAGTG TCTCTTCTGA
3351 GCCTGCCGGA AGACAGGGC TTCCCTACAA TGGAGATGTG CTCCCATGGA
3401 GTCTCTGGCA CTAGTCAGAG AGGGAGAGAG TTTAGGGACT GAAAAACTCA
3451 CCACTGTCA CACCATCACC ATCACCACCA TCATCGCCAT CACCACCACC
3501 ATCACCACCT CAACCATCAT CACTCTCATC ACCATCATCA CTATAATCAA
3551 CACAATCACT ATTGTACCA CCATTACCAC CACCACCACG ACCACAATCA
3601 CTGTTATCAC CATCACCACC ACCCTCAGTC CTCACTACCG TCATCCTCAC
3651 CATCACCGTC ACCACCACCA CCATCACTGC CATCGTCAAC ACCATGGATG

```

FIGURE 3A

```

3701 CTGGTTGTTA AATGCCAGCT CTTTGCCCAA CACTGTCAAG AGTGGTACCT
3751 ACACGGCCTC ATTTTCTGTA ACAACCCCTC GAGGCAAATG TCTGTATCCC
3801 CATTTTACC GAGAGGAGGC CGGGCAGCCT GAAGCACCCG GAGCTGGCAC
3851 TGAGCTCTG CTCTGCATTT GCCACTCCCA GGTGCCTCTG GCCCCAGCTG
3901 GGCCACCTCC AGCACAGGGT GGTGTGTCTT TCCTCAGGAT CTGGGCTCAG
3951 AGCTGCTCTG GGCTGGGGTG CAATCAGTGC CTTGGGCAGG CCCCTCCTCC
4001 TGGGAATGCC TGGTGGCTGA TGCTGGGGTG GGGCTGTGGT CCTTAGGGGG
4051 AGTGTGTGAG CTGTGGGAGC AGCCATGACT GGCTCCCCAG CTGTGCGCAC
4101 AACAGGCCTT CCATCGGTGC CCACAGGTAC TCCTTCCAGG ACGAGGAGGA
4151 CATGTTTCATG GTCGTGGACC TGCTACTGGG CGGGGACCTG CGCTACCACC
4201 TGCAGCAGAA CGTGCAAGTTT TCCGAGGACA CGGTGAGGCT GTACATCTGC
4251 GAGATGGCAC TGGCTCTGGA CTACCTGCGC GGCCAGCACA TCATCCACAG
4301 GTGTGTGCGT GGCAGACGGC GCAGGTACCT GCTGAGGTGG GCGGGGCTGA
4351 AGCAGCCTCA GGTCAAGGCTG CCGGCACGGC GGCGTACTCT CCTCAGAGCG
4401 GGTCTAGCTC CTCTGCCCCA CCCTTGCCTG AGTGCTGCC CCCAGCTGTG
4451 GCACCTGTGC CGACCAAGTC AGCCCCATAG CTGTGTGCTT GGTGTCCATC
4501 TGGGGGGGAC CTGTCCCAGC AGCCCCAGCT GAGACTGGG ACAGTGGGCT
4551 GTTAGCCCTG GTGGACAGAC CACCAGGCTG GGTACAGCA GGTGGCCTTC
4601 ACCTGGTCCA TTTAACTGAA GACTCCTGTT TGCCCATCCA CCACATCCCA
4651 GGGAAATCCAA ACTAATTTTA ACATTAGCTT AAAGCAGATG AAATTAGGAA
4701 GCAGAGCTGG TGTGATGGCT CTGAAAATAA AATTAAAAAA AAGAAAATAG
4751 GAAGCAGATT ATGAAGGAAG TGAATTGGG AAGCAGAAAT TAGGCTGAAA
4801 TTCCCGCAGCA ATGGAACAAA ATGAAAATAT CTGTGAGGTA TATTTTAAAG
4851 TCGAATGGAC TGGTGTTCG ATTTCTGCTC TTGGGGACTC GGATGTCTGA
4901 TTATGACCTA GGCACCAAGT ACTGAGCACT GGCTGTGTAC CTGGAAAAGT
4951 TGGGACAAAG CAAGAGCCGA GGTGGCTTGG TCTCCTAGAG GCCGAGTCTT
5001 GGAGGGGGAG GGCAGACCCT GCCAGCAATT GCTCTCGTCC TCTGGGGCTC
5051 CAGGCCCCCT CCCAGCATCT GGTGCCAGGT GTGTGCTGCT GCCCAGATGC
5101 CACAGGGGGA GAAGTGGCTG ACTTCATCGC CTCTGCCCCC ACGCAGGGGT
5151 GTGAGTCTCT AGCATCATCC AAGGACCAAG TCAAGCTCCC AGGCTCTGCT
5201 CTCGAGTGGG TTGGTGGGAT GTCCTGGGGA CTCCAGGGAT TGTGACAGAG
5251 ATTCAGGGGA AGAAACAGGG CAGATTCCCA ACTCACCTTC CCACCTTCTG
5301 CTCTTTCTAG AGATGTCAAG CCTGACAACA TTCTCCTGGA TGAGAGAGGT
5351 GTGTGGGGTT GGGTGTGGGC AGCCCAAGTG GGTGGTGGCA GGGATGGGCC
5401 TGCAGGGGGA GGAGATCCTT GCACGCAAGG ATGCATCTCT GGTCTGGGGA
5451 CAGCCACACC TGACCCTCTT CTGCACAGGA CATGCACACC TGACCGACTT
5501 CAACATTGCC ACCATCATCA AGGACGGGGA GCGGGCGACG GCATTAGCAG
5551 CACCAAGAGG GTCATGGGTG GAGCCCGAGC TGGGGTTCCA GATGGGAGCT
5601 GGCTTCTCTC AGGTGGGAAG GACAAGACCT CGGTGGCTTC TCTGTCCAC
5651 CCTGGAGGCA GCCTGGTCTC GGGATGTGGC CTCAAGGTGC CGGCCCTGTG
5701 CCCACGGGTC CCGGCTGTGA CCCCCTGGCA GCTGTTTTTC CTCTTTCTG
5751 TCGGAAAGCT CCGGAGATCT TCCACTCTTT TGTCAACGGC GGGACCGGCT
5801 ACTCCTTCGA GGTGGACTGG TGGTGGGTGG GGGTGATGGC CTATGAGCTG
5851 CTGCGAGGAT GGGTATGGAC CCCCTGCAGC CCCCGGGCTT GGCTGCCAGG
5901 CCCCTGCTCT CTGCCCCCAC CAGTGTCTGG GAGGGGGTGG CTGCCCCAGT
5951 GCCCAGGTGC CAGGGGATGT CTCCACTGTG TCTGAGGAGT CAGCCTTTTA
6001 TCGAAGTGTG TAGTTGGTGA TGGAATGCCT GAGCAGGAGG AGGAAGGACA
6051 GACTCACTGT GGTTCCTCCG GGCCTGTGCT GGTGCTGCA GGCAGCCTC
6101 TGTGGGGGTG GACAAGGCTG AGAACTGGCC AGCAGGGGTG CTGCCTCGGA
6151 ACTTTCCACA AAAAGTTTTT TTTGGGGCCC TGTGCTCTTA CCCTTGTGCT
6201 CACGGCGAGG CCAAGTCTGG AGACCGGGAG GCTGGGGGTC CTCTTGTGGA
6251 CCGTACCCCT CAGCCCTGCA CAGGACCCCA CCTCTGAGGA AGCCAGCTCC
6301 CTCCTGGCCC TCTGGGGCTG ATCTACCTGG ACCCAGGCCC CTGGGGATCC
6351 CAGCCAGATG GGCAGCAGC CCAGGGCGCA GGACCCAGGC GTAAGCTTTA
6401 TCTCACCCAG GCTCCTCCGC GGCAGGTGGA GGCCAGGCTG TGCTCAGAGC
6451 TGTGCCTGCA CTTGGGGTGG GGGGAGGGGG TCCTCTCAGG CGCATGGCAC
6501 CTGTGTCTGG CATTGTTCTG GGTGTCTGG GGGCCAGGAG GACCTGCCCA
6551 GCACTGCCTC CCTGTCTCCA GAGGCCCTAT GACATCCACT CCAGCAACGC
6601 CGTGGAGTCC CTGGTGCAGC TGTTCAAGCA CGTGAGCGTC CAGTATGTCC
6651 CCACGTGGTC CAAGGAGATG GTGGCCTTGC TGCGGAAGGT GAGCCCCAT
6701 CCCTGAGCCT CCTCACCTC CGAGCACCCA CCTCCCTCCC TCACTTACCT
6751 GCGGGCTCGC AGACCCCTC CAGTGACACG TTAGTGCCGC TTCCTGGCAG
6801 GCACAGATCC CTTCACTGCA ACCTGTGGGG GCCTCCGCAG ATGGCAGCCC
6851 CAAGCCCCAG GAAGCGAGCT GGTGGCAGGC TCTGTGGCCC TCTCATGGCG
6901 AGCCCTACCA GTGCTGCTGG GTCCTGGTGA TCCCCTGAGC TGCGTCTCCA
6951 GGCACCTCA CAGCAGGCTT GTGCCCCGC CCTGCCGTC ACCACAGAGG
7001 AGCCCTGAGC CAGTCCCCT TGTCACACGT ATTGGAAGGA GGGTGTGGGG
7051 TGGGTGGGTT CAAGCCATG CTATCTCCGG GACCTTTTGC CCCAGTGCCT
7101 CTTGGGGAAG GTGGGAGCC ACTGCCCACT GCAAACACCT CTCAGGGGAC
7151 ATCAGCTGAG CAGGGACACG GGCAGGGGTG TGGCCATCAG TAGTGCCTCC
7201 CATTTGTGAT CTGGTGCTAG GCTGGGGCTG TGCCCTGACT GGGCTGCACA
7251 GCTCCCTGTT CCGAGCTCTG CATTAAACACC ATAGGGGGTT CGGAGTCAGA
7301 GCCAGGCCCA GGCAGGACAG GGAGGGGAGT GAGTGTGCCA CACGGGCCCG
7351 GCTGCCTCCC GGCCCCCGTG TCTCAGGCAG GTGGGGCTCT CTGCCTGGA

```

FIGURE 3B

7401	ATTGTAGCCA	AGCAGCCTAA	AGCCTTGGGG	AGGCCTTGCC	TGCCGGGGCC
7451	TCTCCCCAGC	CCCCAGAGTC	TCTTAACTCT	GCTGTAGCCC	CATGAAGCTC
7501	AGTCACACCT	GCCCAGGTGG	CTCACAAGGT	GGCACTGGGC	TAGAGAGGGC
7551	CTGCGTGGGG	ACTGGGGATG	ACCCACACGC	CCAAGCCCAG	GTCTGGGAAA
7601	CCTCGCACGG	GGTCTGGGTC	TGCGGCATTT	TCCCTGGAAA	GGCGGGAGGT
7651	GCCAGCGCTG	GGATGTTGCT	TCCCAGGCCA	TGCATGGCTG	CCCCGGGCTC
7701	ATCTGGCCTG	TGGAGGTCCC	ATGATTCCGT	GAAGGAAGTG	GCTCTGGGAT
7751	AGTTACTGTG	AGGCCAGCCA	TGTGCCGAGT	GTTAGCCGCT	AGCCGGGCCCT
7801	CGGCTGCCAC	CTCCTGGCAA	ATCCCAGCAG	AGCCTTCCCT	GCAGATCCCT
7851	CTGCTGTCTT	CTGGCGCCAG	GGGTTTAGGT	AGCAGCACTG	AGAACAGGCG
7901	TCCCTTGGGC	CACATGCTGA	GCCAGCCACG	GTGCTTTGCC	TGATGTCGGC
7951	CGTCGGCACC	ACCCCTTCTC	GCGTGGCCCT	GAGGTTCCCT	AATTCTGAAC
8001	CTGAGGCTTG	GTGGGACCCT	CCTCAAGGTG	CCCTGGCCTG	GGGGTGGCGG
8051	GCTATTCCGT	GCTGTGGGCG	TGTGGGCCCT	GGACCCTCTG	ACTCATGCCT
8101	GGTTGCAGCT	CCTCACTGTG	AACCCCGAGC	ACCGGCTCTC	CAGCTCCAG
8151	GACGTGCAGG	CAGCCCCGGC	GCTGGCCGGC	GTGCTGTGGG	ACCACCTGAG
8201	CGAGAAGAGG	GTGGAGCCGG	GCTTCGTGCC	CAACGTAAGC	CTGTGGGCGG
8251	CTCAGGTGGG	GGGCCCTGGG	GATGGATGTG	GCGTCTCCA	CGGGCCGGGG
8301	CTCAGCACCC	ATCCCTCTGT	AGAAAGGCCG	TCTGCACTGC	GACCCACCT
8351	TTGAGCTGGA	GGAGATGATC	CTGGAGTCCA	GGCCCTGCA	CAAGAAGAAG
8401	AAGCGCTGG	CCAAGAACAA	GTCCCGGGAC	AACAGCAGGG	ACAGCTCCCA
8451	TGCCGTGAGT	GCCAGGGCAG	GCTCAGGGCG	CGGCGGCGGG	CTGGGCTTGG
8501	GGCTCCTCTC	TACCACCGAG	CAAGGTGTGT	GGGGACCCCT	GACAGTGAC
8551	ACGTCTCGGA	AGTCCAGCAG	ACCGTTTCCT	GAAGTCCTGA	GAAGGCCAGA
8601	GACCTCCCTT	CTGCCTTTCC	CAGCCCCCAC	CTCGCTCCTT	ATGAAGCAGG
8651	TGGGACAGGA	CAACCAGGGC	TGGGGTTATG	AGTGACACGG	GATGGCCATG
8701	TGAAGCCTTC	GTGCTTGCCC	AGGTGTGCTG	GTGTTGGTTG	TGTGTGCGGG
8751	GACGGCTATG	TGAAGCCCTC	ACACTCGCCC	AGGTGCGTCG	GCATCAGGTA
8801	TGTGTGCGGG	GACAGCCATG	TGAAGCCCTC	ACACTCACCC	AGGTGCGTCG
8851	GCATCAGTTG	TGTGTGTGGG	GACGGCCATG	TGAAGCCCTC	ACACTCGCCC
8901	AGGTGTGCTG	GCTTTGGTTG	TGTGTGCAGG	GATGGCCACA	TGAAGCCCTC
8951	ACACTCGCCC	AGGTGCGTCA	GCATCAGGTG	TGTGTGCGGG	GACGGCCATG
9001	TGAAGCCCTC	ACACTCGCCC	AGGTGCGTTG	ATGTTGTGTG	TGCAGGGATG
9051	GCCATGTGAA	GCCCTCACAC	TCACCCAGGT	GCGTTGATGT	CAGTTGTGTG
9101	TGCAGGGACA	GCCATGTGAA	GCCCTCAGAC	TAGCCAGGT	GTGTCGGTGT
9151	CAGTTGTGTG	TGTGGGGATG	GCCACGTGAA	GCCCTCACAC	TTGCCAGGT
9201	GCGTTGATAT	TAGTTGTGTG	TGCAGGGATG	GCCACGTGAA	GCCCTCACAC
9251	TCACCCAGGT	GCGTTGATGT	CAGTTGTGAG	TGTGCGCAGG	GATGGCCACA
9301	TGAAGCCCTC	AGACTCGCCC	AGGTGTGCTG	GCTTTGGTTG	TGTGTGCAGG
9351	GACGGCCATG	TGAAGCCCTC	ACACTCGCCC	AGGTGCGTCA	GCATCAGTTG
9401	TGTGTGTGGG	GACGGCCATG	TGAAGCCCTC	ACACTCACCC	AGGTGTGTCG
9451	ACATCAGTTG	TGTGTGGGGG	GACGGCCATC	TGAAGCCCTC	ACACTCACCC
9501	AGGTGTGTCG	GTGTCAGTTG	TGTGTGCGGG	GATGGCCACG	TGAAGCCCTC
9551	ACACTTGCCC	AGGTGCGTTG	ATATTAGTTG	TGTGTGCGAG	GATGGCCACG
9601	TGAAGCCCTC	ACACTCACCC	AGGTGCGTTG	ATGTGAGTTG	TGAGTGTGTG
9651	CAGGGATGGC	CACGTGAAGC	CCTCAGACTA	GCCCAGGTGT	GCTGGCCTTG
9701	GTTGTGTGTG	CAGGGACGGC	CATGTGAAGC	CCTCACACTC	GCCCAGGTGC
9751	GTCAGCATCA	GTTGTGTGTG	TGGGGACGGC	CATGTGAAGC	CCTCACACTC
9801	GCCCAGGTGC	GTCAGCATCA	GTTGTGTGTG	TGGGGATGGC	CACGTGAAGC
9851	CCTCAGACTA	GCCCAGGTGC	GTCGGCATCA	GGTGTGTGTG	CCGGACAGC
9901	CACGTGAAGC	CCTCACACTC	GCCCAGGTGT	GCCGGCTTTG	GTTGTGTGTG
9951	CGGGGACGGC	CACGTGAAGC	CCTCATGTCT	ACTCAGGCAT	GCTGGTATTC
10001	TGGGGCTGCC	AGGACAGGTG	ACCACGAATC	AGGTGTTGA	AGAACAGCAA
10051	TGCGTCTCTC	TGAGAGGATC	TGAGTCGTAA	TGAAATGGTC	TCCTTCACAG
10101	CCGGCTGTGC	GTGAACTACT	CTGTCTCCTG	CAGCTCCCCT	GTCTTGATAA
10151	TTGGCTGTCT	AGGCAGCGGG	TAAGGTGAAC	CCCTTGGGCA	GTTATGTGAT
10201	GATCTCAGTT	TCTGTAAACC	GGAAGTCCAG	GCATGGTGCA	GCTCTGTTCC
10251	CTGCTTCGGG	GTCTCACCAG	AATGTGAGCT	AACATTGAGG	TCGTGGCCTT
10301	GTCAAGTGCA	GCTCTGTTTC	CTGCTTCTGG	GTCTCCCCAG	AGTGTGAGCT
10351	AATATTGTCT	GAGGTCGTGG	TCTCATCAGG	GATTTGACAG	GTGCTGTGGT
10401	TGAAATGTTT	CCCTTAAAAC	TCGTGTTGGA	ATTTGCTTCC	TATTGTGATG
10451	GTGGTAGAGA	TGGGACTTTT	GGGGGCTGAT	GGGGCCACGT	AGGTTCTTCC
10501	AGCATGGATG	GGGTTAATGC	TGTTGTAGAA	GGGTGACTTT	AGTCCTCTTT
10551	TGAGTCTTTG	ATCCTCTGCT	ATGTGAGGAC	GTGGTGTTCC	CAATGTGGAC
10601	GTGGTTTCGT	TTCCATGTGA	ATGTGATATT	CACAATAGAG	CATCAACAGG
10651	CTCCCTTTTA	ATCAGCAGAT	TTAAAAAGAA	ATGTGTTGTC	TCATGGCTTG
10701	GAGGCTTGAG	TCCAAAGTTA	AGATGTCAGC	AAAGCCGTGC	CCCTCTGAA
10751	GGCTCTCCGG	GGAGGAAAAC	CAGTCTTTGC	CCCTCTACCC	TCCGGTAGAG
10801	GCTGCCTTGG	CCTAGACGCA	TCCCCCAGC	CCCTGCTTCG	CTGCCGCGTG
10851	GGGTGCGCCT	GTGTGTGCGT	CTCCATCTCC	TCCCTCTTTC	TCATAAGGAC
10901	ACCAGGCATT	GGATTTAGGG	CCCACCCTGA	TCCAGTATGG	CCCCATCTTA
10951	TCTTGATGAT	ATCTGCAAAG	ACCTCACTTC	CAAATGAGGT	CACATTCA
11001	GGTACCCAGG	ATTAGAATTT	GAGTGTGTCA	TTTTTGGGGA	CACAGTTTGG
11051	CCCATACCCAC	CAGGATGTGG	CTGATATTCA	CCAAGGAGTA	GCTATGGTTG

FIGURE 3C

11101	TGTGTTGATG	TCAGGGTGAC	GGTGATGACC	CTGGGTCCTT	CGGTGGTCCC
11151	CTTGCCCTCG	AGTCTGCCTG	AGCCTGTGGT	GGATGTCCTG	GGAAACTCTT
11201	GTGCCTCAGC	CCCCGTGCAG	CCTCCTCAGA	CCTGGTGGGC	CCTGTGTTGC
11251	TCCTGGGCAG	AAGACGGGTG	TCAGTCCCCT	CCTCACCATG	ATGTGGGGGG
11301	CAGGGGTGGG	GTCAATGCCCT	GGGTGCCCTG	ATTTTGGGGG	GAACACGGCC
11351	CCCCAGTGGG	TCAGGCTCCC	ATCCTCGCCC	CTCCTCCAGG	ACGGCTGCCG
11401	GCAGCCCTGG	GTGTCTCTAG	GCAGAATTGC	TGGTGGAGAG	CTGCTGTCTG
11451	CCAGGTGGCC	ACTGTGAGGC	ACTGCTGAGA	GCCACAGGAT	GGTTGGAAGG
11501	TTCTCGGGGT	TGGGGGTTCT	TTGGCATTGC	CCCCATTGGA	TGTTTAAAGT
11551	TTCCCTACCA	GAGCATGTCC	AGAGCCAGGG	CTCTGGGGTG	TAGAAACAGG
11601	CCCAGGATGA	GTTAGGAACC	CTCATGGGAG	ACTCAGGGAT	GGACAGTGTG
11651	CAGAGCCAGG	CTGGCCATGC	TGAGTTCCCA	GGAGGCTCTG	GCTGGGAACA
11701	GGTAAGGCCA	GGCACCTGTG	AGCGGGAGGA	GCTCGGCTTT	GTCTTGGGTT
11751	GCTTGTGTGG	AGATGTTTTG	GCTTGAGGGT	AGGAGGTGTT	CTGAAAGGAA
11801	AGCATCACTC	CAAAAAAATA	GTCCCACTGT	TAACCTTGAG	GCTGAGAGAG
11851	GTTTTTGGA	ACAGCTTTAT	TTTGATATAA	TTCAATTCC	ATGCAATATA
11901	CAGTGCATCC	ATGTAAAGCA	TATAATTCCA	TGGTTTTTAA	TATAGTCACA
11951	GGGCTGTGCA	TTCTCCACCA	CAATCTGATT	TTAGAACCTT	TTCATGTAAT
12001	GTAAGAGAAA	GACCCACCT	ATTAGCAGTC	ATGCCCCATT	CCCCTCTTCT
12051	CCCCTCCCCT	GGCAGCCACG	AAGCTACTTT	CCGTCTCTGT	AGGGTTGCCT
12101	GTTGTGGGCG	TTTCATGGAA	GTGGAGTTAC	ACACTATGTG	GTCTTTGCAG
12151	CTGGCTTCTT	TCACCTCGCA	GGATGCTTTT	GAGGCCCGTC	CACGTTGTAG
12201	CCTGTCACTG	CTTCATTCTT	GTTGATGGCT	GAGTAATATT	CCACATATGT
12251	ATCACCTTTC	CTTTATCCAG	TCATCAGTTG	ATGAGTATTT	GTCTTTTCCA
12301	CTTGTTAACA	TTTTTCATTA	TCATGAATAA	TGCTGCTGTG	AACATTCAAC
12351	TACAAGTCTT	TGTGTGGATA	TGTACTTTTA	TTTTTGGGGG	GCACATACTC
12401	AGGCATGAAC	CCGCTGGGTC	ATATGTGACT	CTGTGCTTCA	TGTTTGGAGG
12451	AAACACCTAC	CCTTTTCTAC	AGCAGGTGTG	CCATTTTATG	CCCCTACCAG
12501	CAGTGTGTGA	GGGTCTCAAT	TTCTCCATAT	ATTTACCAAG	TCCTGTTATT
12551	GTCTGGTTTA	TTTTTTTAAA	AATCATAGTT	ATCTTAGTGT	GCAGTGATTG
12601	TGTGGTTATG	GTTTGCATTT	CTCTGATGAT	ATTGAACATC	TTTTGAGGTG
12651	TTTTATCAGG	CATTGTGTCT	AGAGAAATGT	CTATCCAAAT	TTTTTAAAT
12701	TTTTATTGTC	TTTTTTTTAG	TCTACTCTGA	CAATATATTT	TAATTGGCAT
12751	ATTTATTTTA	CTTTATTTTT	TTTTAGAGAC	AGGGTCTTGC	TTTATTACCA
12801	AAACTGGAGT	GAGTGGGTGC	AATGAAGGCT	CACCTCAGCC	TTGACTTCCT
12851	GGGCTCAAGT	GATCCTCCCT	GCCCCAGCTG	CCAGAATGGC	TGGAAGTGTG
12901	GGTGTGCACC	ACCACACCTG	GCTCATTTGA	AAAAAATTTT	GTTGTAGAGA
12951	CAGGGTCTCA	CTATGTTGTC	CAGGTTGGTC	TCAAACCTCT	GGAGTCCTCC
13001	CACCTCAGCC	TCCCAAAATG	CTGGGATTAC	AGATGTGAGC	CACGTGTCCT
13051	GACCTAATTG	GTGTATTTTA	GACCATTAC	ATTTAAAGCG	ACCAGGGAGG
13101	CTGAGGCAAG	AGGACTGCTT	GAGTCCAGGA	GTTTGAGACC	AGCCTGGGCA
13151	ACAAGGTGAG	ACCCATCTGT	ATTAGTCTAT	TTTCACACTG	CTAATAAAGG
13201	CATACCTGAG	TCTGGGTAAT	TTATACAGGA	AAAAGGTTTA	ATGGATTTAC
13251	AGTTCCACAT	CTGTGGGGAG	GCCTCACAGT	CGTGAAGGCG	AAGGAGGAGA
13301	AAGTCACATC	TTACATGGAT	GGCGGCAGGC	AAAGAGAGAG	CTTGTTACAGG
13351	GAAACTTTTG	TTTTTAAAC	CATCGGATCT	CATGAGACTC	ATTCATATC
13401	ATGACAACAG	CACAGGAAAG	ACCCGCCCCC	ATAATTCAAT	CACCTCCAC
13451	CAGGTTCTCT	CCACAACATG	TGGGAATTGT	GGGAGTCACA	ATTCAAGCTG
13501	AGATTGGGAT	GGGGACAGAG	CCAAACCATA	TCATTCTGCC	CCAGCCCCCT
13551	CAAAATCTCA	TGTCCTCACA	TTTCAGAAC	AATCATGCCT	TCCCAACAGT
13601	CCCCCATAGT	CTTATTTTGG	CATTAACTCA	AAAGTCCACA	GTCCAATGTC
13651	TCATCTGAGA	CAAGGTAAGT	CCCTTCTGCC	TATGAGTCTG	TAAATCAAA
13701	AGCAAGTGAC	TTCTAGATA	CAATAGGGGT	ACAGGCATTG	GGTAAATTCA
13751	GCCATTCCAA	ATGGGAGAAA	TTGGCCAAAA	CAAAGGGGCT	ACAGGCCCCA
13801	TGCAAGTCTG	AAATCCAGCA	GGCCTGTCAA	ATCTTAAAGC	TTCAAAATGA
13851	ACATCTTTGA	CTCTATCTCT	CACATCCAGG	TCATGCTGAT	GCAAGAGTTG
13901	GGTTCCCATG	GTCCTTAGGCA	GCTCTGCCCT	TGTGGCTCTG	CAGAGTACAG
13951	CCTTCTCTCC	GGCTGCTTTC	GTGGGCTGGC	ATTGAGTGTG	TGTGGCTTTT
14001	CCAGGTGCAT	GGTGCAAGCT	GTTGGTGGAT	ATACCATTTT	GGGGTGTAGA
14051	GGATGGTGCC	CCTCTTCTCA	GAGCTCCTCT	AGGCAGTGCC	CCAGTGGGGA
14101	CTTTGTATAG	GGGCACCAAC	CCCACATTTT	CCTTCTGCAT	TGCCCTAGCA
14151	GAGGTTCTCC	ATGAGGGCCC	CACCCCTGCA	ACAAACTTCT	GCCTGGACAT
14201	CCAGGTGTTT	CCATACATCC	TCTGAAATGC	AGGCAGAGGC	TCCCAAACCT
14251	CAATTCTTGA	CTTCTGTGCA	CCTGCGGGCT	CAACACCACA	TGGAAGCTGC
14301	CACAGCTTGG	GGCTTGTACC	CTCTGAAGCC	ACAGCCTGAG	CTGTACCTTG
14351	GCCCCTTTCA	GTCATGGCTG	GAGCAGCTGG	GATGCAAGTC	AGCAAGTCCC
14401	TAGACTGCAC	ACAGCAGAGG	GACCCTGGAC	CTCGCCCATG	AAACCATTTT
14451	TTCTCTCTAG	CCCTCTGAGT	CTGTGATGGT	AGGGGCTGCC	GCAAGGCTCT
14501	GTGGCATGCC	CTGGAGACAT	TTTCCCCATT	GTCTTGGTGA	TTAATATTCA
14551	GTTCTTGTGT	GCTTATGCAA	ATTTCTCCTC	AGAAATGGG	GTTTTCTTTT
14601	TTTTCTCTCT	TTTTTTTTTT	TTTTTGAGAC	AGTCTTGCTC	GTGCACCCAG
14651	GCTGGAGTGC	AGTGGTGCAA	TGGCGGCTCA	TTGCCACTGC	AACCTCCGCT
14701	TCCTAAGTTC	AAGTGATTCT	CCTGTCTCAG	CCTCCCAAGT	AGCTGGGATT
14751	ACAGGCACGC	ACCACCACAC	CCAGCTAATT	TTTTGTATTT	TAGTAGAGAA

FIGURE 3D

```

14801 GGGTTTCACC ATGTTGGCCA GGCTGGTCTT GAACTCCTGA CCTCAGGTGA
14851 TCTGCCCTGCC TTAACCTCCC AAAGTGCTAG GATTACAGGC GTGAGCCACC
14901 GTGCCCGACCC AGGAGTTTCT TTTCTATTGC ATTGTCAGGT TGCAAATTTT
14951 TTGAACCTTT ATGCTGTTTC CTTTTTAAAA TGGAAATGCCT TTAACAGCAC
15001 CCAAGTCACC TCTTGAATGC TTTGCTGCTT AGAAATTTCT TCTGCCACAT
15051 ACCCTAAATC ATCTCTCAAA TTCAGAGTTC CACAAATCTC TAGGGCAGGG
15101 GCAAAATGCT GCCAGTCTCT TTGCTTAAAG CATAACAAGA GCCACCTTTG
15151 CTGTAGTTCC CAACAAGTTC CTCATCTCCA TCTGAGACCA ACTCAGCCTG
15201 GACTTCATTG TCCATATCAT TATCAGCATT TTGGTCAAAG CCATTCAACA
15251 AGTCTCTAGG AAGTTCCAAA CTTTCCCACA TTTTCTGTCT TTCTTCTGAG
15301 CCCTCCAGAT GGTTCCAGCC TCTGCCTATT ACCCAGTTCT AAAAAGTTGC
15351 TTCCACATTT TCAGGTATCA TTTCAGCAGC GCCCTACTTT ACTGGTACCA
15401 ATTTACTGTA TTAGTCTGTT CTCACGCTGC TAATAAAGAC ATATCCGAGA
15451 CTGGGAAATT TATACAGGAA AAAGGTTTAA TGGACTTACA GTTCCACATG
15501 GCTGGGGAGG CCTCACAATC ATGGCGGAAG GCAAGGAGGA GCAAGTCACA
15551 TCTTACATGG ATGGCAGAGA GAGCTTGTGC AGGGAACTTT TTGTTTTTAA
15601 AACCATCAGA TCTCATGAGA CTCATTCACT ATCATGACAA CAGCACAGGA
15651 AAGACCCGCC CCCATAATTC AATCACCTCC CACTGGGTTC CTCCCATGAC
15701 ACACGGGAAT TGTGGGAGTC ACAATTCAAG CTGAGATTGG GGTGGGGAGA
15751 CAGCCAAACC TTATCACCAG CTCTATAAAA GACAAAAAAA TTAGGCAGGC
15801 ATAACAGTGC ATGCCTGTAG TTCCAGTGAT GTGAGAGGAT TGCTTGAGTC
15851 CAGGAGTTTG AGACCAGGCT GGGCAACATG GCGAGACCCT GTCTCTACAA
15901 AAAAAAATTA TCTGGGTGTG GTGGGATACA CCTGTGATAC CAGCTACGCA
15951 GGAGGCTGAG GCAGTAGGAT TGCTTGAGCC CAGGAGTTCA AGGCTGCAGT
16001 AAGCTATGAT CATGCCCTTG CACTCCAGTC TGGGTAACAG AGAGACACGC
16051 TGTCTGTAA ATAAATAAGT GGTCAATTTAT ATAGTTCAAT ATGATATCTA
16101 CCTTATTGTG AACTGTAGTC TATTTATTGG TCTTCTTTT TCCCTATTTT
16151 TCTGCCTTTT CTGGTTTTAA TTAAGCATTT TATATTATTC TAGTTTATCT
16201 CCTCTCTGG CCTGTTAATT ATACTTCTTT TTGAAATATT TTTAGTGGTT
16251 GGCCTGGACA TTGCAGTATA CCTTACCATT ACAGTCTACC TTCACCTGAC
16301 ACTCTGCCCC CTCATGTGCA GTGGGATGCC TTGTGACGGC ACCTCTCGTG
16351 CAATCCCTCT GTTCTGTATG ACATTGCTGT CATTCAATTT ATTTATCTGT
16401 ATGCTATAAT TGCTCATTAC ATTGTTACTA CTGTTATTTT AAACAGTTAT
16451 CTTTGGATC AATTAAGAAA AATTAAAAAT TTCATTTTAC CTCTATTCAT
16501 TCCTCTCTA AAGTGCTTCC TTTCTTTATG CAGACCCAAG TTGCTGACCT
16551 AAATCATTTT CCTTCCCCTT GAGGAACTTC GTTTAACATG TCTTATAGGA
16601 CAGGTCCAAC AGAGATGAAT TCCTTCCCTT TTTGTTTGTC CAAAAAAGTC
16651 TTATCTTTCA CCTTTAAAGA ATAATTTTCA TGGATATAGA ATTCTAGATT
16701 GGTAGGTTTT TTACTTTCAA CACTTTAAAT ATTTCACTCC GCTCTCTTCT
16751 TGCTTATGTG ATTTCTAACA GGCAGTCTGC TCTAATTCTT TTTCTGTAAG
16801 TAGAATTCCT CGCCACCCCC CACCCCAAGC TCATTTCAAG ATTTTCTCTG
16851 TCTTTGGTTT TCTGCAATTT GAATAATGAT ATGCCTAGGT ACAGATTTTT
16901 TTTTAATATT CATTCTGCTT GGTGTTCTCT GAGCTTCTTA GATCTGTGGT
16951 TTGGTGTCTG TATCTAATTT CAGAAAATTC CCAGTCATTA TTATTTCAAA
17001 TATTCTGTTC TCTTCTCTTT CTCCTGCTGG AAATCCAATT ATGTGTGTGT
17051 GGTGCCGTTT GAAATTGTCC CACAGCTCTT GGATATTCTG TTCATTTTTT
17101 TCACTCTTTT TTTTTCTCT TTGCAATTTCA GTTCGAGAAG TTTCTGTTGA
17151 CATTTCTTCA AGCTCATAGA TCCTTCTCTA TCTGGATCCA GTCTACTGAG
17201 GAACCATCAA AGGCATCCTC ATTTCTGTTA CAGTGTTTTT TACTTCCAGC
17251 ATTTCTTTTT GATCCCTCCT TGGAGTTTCC ATCTCTCTAC TTACATTGCC
17301 CTTCTGTTCC CCATCTGTTT TTGCCTGCTT TCCCCTGGA GCCCTTGCCA
17351 TATTAATCAC AGTTATTTCA AATTTCTGT TTGATAATTC CAACATGGGT
17401 GCCATATCTG GATCTGTTCT AATGTGTGCA TTGTCTCTTC ACACTCTATT
17451 TTTTTCTTTA AGACCTTGTA GTCTGCCCTG TAATGTTTGT TGAAAGCTGG
17501 ACGTGATGTA TCTGGTAATA AGAGCTGAAG TAGATGGGCC TTTAGTGTGA
17551 GGTTTATGTA ATCTGGCCAG GTTTGGGGCA GGGTTTAAAG TCTGCTGTAG
17601 CTGTGGGTAC CAGAGGCTTC ACATTTGTTT TCATTTCCCG GGTGGCCCTT
17651 GGGCTTGCCCT AAATCCTCCT CCTCAGAGAG AGTCTGCGTC TTGTGGCCCT
17701 CTCTGCTGGA ATCCCTGTCA CACTGCGGAG GCCCTGTGGG TGTTTGGGA
17751 AGATGTTGGG GAGGGAACTG TTCCACAGTC TGTGACCAAA TCTCAGTCTT
17801 GGGGGCTGTG GCCCTTTCAC AGTTGTTGAT CTGCTTTTCC CCTCCCCTTA
17851 GGTGAGACAG GCTAAAGCGG GGTACAGGCT GAAAAACAGT CTTCCCCCAA
17901 GTGAGATAAG CCTTTCCTTT GGAGAGCAAA TTCCATTTGC TGTGGAGAAT
17951 GCTCTGGGTG TATTTACAG TGGTGACTGT CCCCATCCCA TGCCAGAGCC
18001 AGGAAGGGAT CATCCTTGGC TTCATTAAGA CCTGGCAGGG TTCTGGAGG
18051 GGAAACCCAC AAAACGTTGG GGGCCTCATA AGACCGCAGC CGCAGGAGTT
18101 CTGCACACGG CCCCAGCCGC TCCCAGAGC TGCCAGGAG GTGTTCCCGC
18151 ACACCATCAG TTCTGCTCCA GGCCAGCAGG TCTCAGCTGT GACTTTGCTC
18201 ACCTGTCTCT CCAGACTTGG GGGTCGCCGC AGCGCTTTGA CCTCAGCTCT
18251 CTGATGGGTC CAGGAAAAGT AATTGATTTT CATTAGTTCA GCATTTTTC
18301 TGTCGACAGG ATGGGAGTGA GGCCTTCCAC GTTGTGTTGC GCAGAGCAGA
18351 AGCCAGAAGT CCTTGCTCGT CTCTGTGTTT CTTGCTCGGA GTCATGTCTT
18401 TTATTTCTCT AACGCAGGTG TGTGGTCAGA CAGGAGATTG GCAGGTATTT
18451 CCTCTGTCTC CATGGGCTAT CTTTTCTCTT TCTTGAGGCG ATCATTTGCA

```

FIGURE 3E

18501	GCAGGGAAAGT	CTTGGACCTT	GACGTTGTCC	GATGTGGCTG	TTTCTCCTT
18551	GGCTGCTCTG	CTTTGTCTAA	GGATCCATCA	CCTCAGCTGA	GGCCACAGGG
18601	ATTTACTCCC	ACACTTTCTT	CTAAGTCTTG	TATAGTTTCA	GCTCTTGCCT
18651	TTAGCTGTGT	GATTCAATTT	GGGTAAATTT	TTATGAACAA	TGTCAGGTGA
18701	GGGTCCAGCT	TCATTCTCTC	ATCGGTGGAT	ATCTGACTGT	CCTAGCACCA
18751	TTTGCTGAAG	AGAGGATTCT	TTCCCATTTG	AATTGCTTTT	GACATCATAC
18801	CTTGTTTTTT	GACTTGCCGT	TTTATCCCAT	TGGTCCAGAC	GTCTGCCCTG
18851	CGCCGGGACT	GCACCATCTT	GATAACTGTA	GCTTTGTAGC	AAGCTTTCAA
18901	ATCAGGATCT	GTGAATCCCC	CAGTTTTGTT	CTTTTTGGAC	ATTATTTGAC
18951	TATTCTGGGT	CCCTGGCATT	TCCACTGAAT	GCTGAGGGGT	CTGACAGTTG
19001	CATCTGAGCT	GCCAAGCAGG	TTTGTGGCGG	TGCTAGGGAC	TGAAGCCTGC
19051	TCCATTTCCC	AGGCCCTCC	TCGCTGTGGG	TGACATCTGG	GGTCCGAGGC
19101	TGTGTCTCAG	CATGTGTGAA	GGTGCCACGG	GTGCCCTGAG	ATGGGGATTG
19151	CTGGTCCAGT	TACTCAGAAA	GTGCATCCAG	GAGAGACCCC	CGCCCTTCTC
19201	GGGATGGGAG	ATGCCAGCAG	AGCTTGGCTT	TCAAGCAGAA	ATCTGGAAC
19251	CCTGTGGGGA	GTGGCTTCGG	ACTTCAGGGG	ACCTGGAGCG	TCATTTGTGG
19301	TTCAAGAGGT	CCCTGCCCTG	AGGGAGCTGG	GCTCTCAAAC	ACCCACACCA
19351	GTCAGTCATG	GGCGAAGGGC	CCCTCTGTGG	CCTTCTGACT	GTGTGTGTGC
19401	TGGCAAAGGG	TTCCAGCCAC	CCAAGGAGGA	GGCAGGGGCT	GTGAGAGGAA
19451	GAGCAGGGCA	GATCCAGGCC	AGGCACAGAC	CCTGCCATGG	GGTACTGCTG
19501	GCGGCCTAC	ACACACCTTG	GGAAGGGAGT	CCCTGTGGAA	AGGGGTCTGT
19551	GTACACATCT	TAGGTGACAC	AGCCCGGCTT	GGGCGCTGCT	CAGAGCCACC
19601	CCTTCCAGAT	GGTTCTGGAG	CAGCTCCTCA	GGCTTCTGGT	GGCCTCTCTG
19651	CCTAGGAAAA	CATGGCTGTG	GACGTTGCAG	GATGACCAAC	AGCCCTGCC
19701	ACTGGGCTGC	ACAGAGGGCC	ACGACGGGCG	CTCATGTTCT	ACATCACTGG
19751	CGCCCAACCC	AGCCCTCCCT	ACCTTGTGTC	GCTGTGAATC	GCAGGATCCC
19801	AGCGGCTCAG	TCGGACCCTC	ATTCTGAGT	AGTCTGAGCC	TTAGGTCACC
19851	GTACCTTCT	CAGGCCGGCC	CGAGTTTGCA	GACTTGTCTG	TCTATATCAG
19901	GGTTAGACCA	GAGAGTGCTG	AGACACAGCA	GATCACCCAG	CCTGTCTCT
19951	TCTTGATGAC	TAAGGACAGG	TCCCCTGCCA	GGATCGTGAC	TCCTTTAGGG
20001	GAGGCCACAG	TGACAGGGCA	AAGCCTGGAG	GGAGAGAGCC	ACATGGAGAG
20051	GAGAGGCTG	CCCGCAGAGA	GCGTGGGAGT	CTGCCGGCTT	CTTCTCGAGT
20101	CCTTGCCAAG	GTGCTGGCCG	CTCACACCGT	GTACGTGTGG	GGAACGCCCA
20151	GGACCAAGGT	GACACCACCA	GGAGGAGCGG	GCGGACAGT	CCCCACTCAG
20201	GGCTAGGAAG	AGAGAGTCCA	CATTCCCCTG	GCCAGGGTGA	AACCCTCACA
20251	CCACCACAGA	TCCAGGAGAG	ACACGGAGGG	CACTGCCTCG	GGGTGGGGAA
20301	CGTGAGCTGC	TCCCAAAACC	CAAGAAATGT	GTTGAGCCCT	ATGCTTCCTT
20351	CCTCGTGGGA	AGAGGCGCAA	GGTGAGACCG	CTTGTCTTTT	ATCTTGGCGG
20401	GCACGTCCGA	GTGTGACGCA	CGTCAGCAAA	TCCCTGAACC	GTTTCATCGGA
20451	GAACAGCCTT	CTGCATCTCC	CACACTCTGT	TCGTGGGTTT	ACAGGGGTGC
20501	CAGAGTACTT	GACAGTTGGC	AATCAGCATT	AATAGGATCC	ACAGGGCCAG
20551	GCATGTTGGC	TCACGCCTGT	AATCTCAGCA	CTTTGGGAGG	CCGAAGCAGG
20601	CGGAGCAGCT	GAGGTGAGGA	GTTCAAGACT	GGCGTGGCCA	ACATGGTGAA
20651	ACCCCGTCTC	TACAAAAATT	AGCTGGGTAT	GGTGGTGGGC	ACCTGTAATC
20701	CCAGCTACTT	GGGAGGCTGA	GGCAGCAGAA	TCACTTGAAC	CTGGGAGGCG
20751	GAGTTGGCAG	TGAGCCAAGA	TCACGCCACT	GCACTCTAGT	CTGGGTGACA
20801	GAGCGAGACC	AAAAATAGGA	TCCATGGATA	GCAGGCAAGA	GTGTCCAGGT
20851	GTTGAGGCA	CAGACGACAC	TGTGACAGGG	AAGAGTCCCC	TTAGCCCTGG
20901	CTGGGGCCGT	GAAAGCATGC	TGTTGTCCGT	CTCGGGTGAA	CGCAGACTGT
20951	TGTACAGCAT	TGCATAACGA	TGCTTCCGTC	ACTGGCCAAT	CGCATGGGGG
21001	GGTGGTCCCG	TAAGATGGTA	ACACTGGGTT	TTGCTGTACG	TTTTGTATGT
21051	CTAGATAGGG	TTGAGCGTTC	TGGTGTGTAC	CCACTCACAC	GTCCCTCCCG
21101	ACCTTCAGAG	CCCAGCTCCC	TCCCTCCAG	GGCCTTGGCT	GTGACGTGGG
21151	TGACTTCCTA	TGGATCTGAG	GTTCTGTGGT	CCTCACAAGT	GGGCATCCTC
21201	TGGCCTCAGC	TGCAGGAGGT	GGGGGCCCTT	TTAATGCCAC	CCGAGGCCTG
21251	CGACTCCCTG	CACTTTTCAC	TGTGACTTGG	CTCATCTGGG	TCTGTCATTT
21301	GCTCACGCGT	TGGTAGTGAC	CAACGTACCC	ATCCAAGTTC	ACGGTCACCA
21351	TAATGATGCT	TTCCCCACAC	CATGCCAGCG	CTGAGCGGCC	AGCACCCCTT
21401	CCACCCACCC	CACGCCCTCC	CCCGACCCCT	GCCTCCTGGG	AAGTGGTCCT
21451	GCTGCCTGAA	GGAAGTAGTC	CTGCCTGCCC	ATCACACACC	AGTAAGGGTG
21501	GGTCCTGCCA	GGGGCAGCCT	CCGTCCACAA	GCTTGCCCTG	AGGACCTGCT
21551	TCTAAGACAG	CCCTGGTTCC	AGGATTCTCT	GGGCAGGGCC	CCAGAAGCAG
21601	GCCTGGGACA	GGTGTGTGTG	TGCTGTGATG	AGGGGCTGGG	AGAACCCGGT
21651	ATGAGACGGG	AAAGGCCCGG	CAAGGGAGTG	GTTTCCAGCA	AAGTCCCGCA
21701	GAGAACAGCT	TCTGCCTGGT	CCTGCAGGCC	CCACGGAGCA	AGTCCAAGCC
21751	ATCCACCCAG	AGGCAAGGGA	GCTGGGCCTT	GGCATCCTCG	GGCTTGGGTG
21801	AGTCACCCCG	AGAGATGCGA	GCTCCCCGGG	CAGTCTGGCT	GCTGGAGGGC
21851	CGGGGCACCT	CAAAATAGCCC	AGAGGCCGTC	ATCCAAAGCC	ACAGGTGGAG
21901	GCCCCATGGG	GATGCCCAGA	CACTCACTTG	AGGGGACATG	GGCGGAACCT
21951	GGACAGCGTC	CCCCACGCTC	ACGTGTGCCT	TTCCATCCAC	AGGAGAATGA
22001	CTATCTTCAA	GACTGCCTCG	ATGCCATCCA	GCAAGACTTC	GTGATTTTTA
22051	ACAGAGAAAA	GTGAGTGTGT	TGGGGGTGGG	TGGGCGTGGT	GGCAGAGAGG
22101	AGGAAATAGG	GGCTAAGGTT	AAGGTTTTCT	TGGCCACGTG	AGCGGGCACC
22151	TGTGGGCTCG	GGGTGCGTGG	CCCTGCTCTC	TTTGGGGACT	CTGAGCAGCA

FIGURE 3F

22201 GCTATGGAGG GGAGCGGCGG GAGGCCCTG CCAGGCTCTG GCATGTTTGT
22251 GCTCCACGCG GGGCCCGTGG CTGGAATCTT CTGGGGAGAG ACACATCATT
22301 TGCCACAGATG AGGGGTGGTG ACTTCCTAGG AGGCCCCATC AGAGCCACGT
22351 CAACTCCCCC ACCCAGGCAC GCCCTCAGTC TCTCAGCAGA CCTTTCCTGA
22401 ATGTACAGGCC CCAGGGGACA GAAAGGGCAC AGATGACTGG CAGCAGGCAA
22451 GGCAGGCCAG AAATAGCAGC AGCTGCCACG GTGGGGCCCA AGGGAGGATG
22501 GATGCTCCCT CTGCCCACAC GGGGCAAGGA GGGCCTCCTG GAGGAGGTGG
22551 GTCTGAGCTC TTATGGACAG GACGTGCAGG GCAGCACGTG CAGACGGCTG
22601 AGGGCACTGA CTGGCACCTT GGGGATCAGA CGACCGGGTG AAGAATGAGG
22651 CTTAGCCGAG CCTCATTCCC AAGTCACTGA CCTATGGCAC CTGCACAGTC
22701 AGGCCTTTTC GCTTCTGGCT GGAAACATGC CGAGCCTCGC CAGCATGCTC
22751 ACGTGCCCCC ACCCGTCCCC AGGCTCCCTG CCAGTGTGTC GGGAGCATGG
22801 CCTCTCCAGC AGACACCGAG CCTGTGGCCC ACGTTTGGGC ATCCACGCCA
22851 TGGCCTATTC GCTGAGCCCG CTGAGCCCGT TGGGCAAGTC ATGGGACCGT GAGGCCAGGG
22901 AGGTGGGGGG ATAACGCCCT CCATGTGTTC CTGCCACCCC AGGCTGAAGA
22951 GGAGCCAGGA CCTCCCGAGG GAGCCTCTCC CCGCCCCCTG GTCCAGGGAT
23001 GCTGCGAGC CTGTGGAGGA CGAGCGGAA CGCTCCGCCC TGCCCATGTG
23051 CGGCCCCATT TGCCCCCTCG CCGGGAGCGG CTAGGCCGGG ACGCCCGTGG
23101 TCCTCACCCC TTGAGCTGCT TTGGAGACTC GGCTGCCAGA GGGAGGGCCA
23151 TGGCCCGAGG CTTGCGATTG ACGTTCACAC CCAGCCTGGC TGCGGTGCC
23201 CACAGTCCCC CGGACACATT TCACACCTCA GGCTCGTGGT GGTGCAGGGG
23251 ACAAGAGGCT GTGGGTGCAG GGGACACCTG TGGAGGGCAT TTCCCGTGGG
23301 CCCCCGAGAC CCGCCTAGAT GGAGGAAGCG CTGCTGGGCG CCCTCTTACC
23351 GCTCACGGGG AGCTGGGGCC ATGGATGGGA CAGGAGTCTT TGTCCCTGCT
23401 CAGCCCCGAG GCTGTGCACG GCCCTCGTCA CAAGGTGACC CTTGCAGCAC
23451 AGGCCCGCGG TCGCCAGGTC TCGGCTCAGT TCTTGAGGTT CAAGGGCATG
23501 GGTGGGGTGA GTGGGTGGGG AGGTGAATGT TTTCTAGAGA TTCAAACCTG
23551 TCCAGCAATT TCTGTATAGT TTTCACTCTT GAGAATTACA ATGTGAGAAC
23601 GCGTACGTCC TGCATGTTCT GCGTACGTCC TGTGTCTGCC TGGCCGTGAG
23651 GCGGCTGCTC GCGGTTTCTG GTTGGCTGAG ACTTGGGGCA GCCAGTGGGG
23701 TGGGCACTCT CTCAGGGCAG AGCTCCCGGA CCATGGCTTT GGGGTGGGTG
23751 CTTCTCTCCG TGGCCCTGGA GCCGTAAGGC TGTGGAAGGC AGAGACGGTC
23801 CTGGAGGCAG AGGAGCCAGG GACAGCACCG TGCACCGTGG AGCCGCCGCA
23851 GTGCCGGGCA GTGCTTGGCC CTCCATAAAG GGACGTATCC CTCTCACTGT
23901 GCGTGGGTGG TTCTGTGGTT GGAACGTGTA CTAACCTGGT AAACGGCCTG
23951 TGTGCTTCTC TCTGGTCTCG CTGGAGGAGG ACGGGCTCAG CCCGTGAGCC
24001 CAGCGCTCCA GACAGGCGCT TGCTGGTTTC CTCTGAGGAA ATGGGTGTGG
24051 CGGGCTGTGT CCCCTTCCCA GGACAGCGGC CATAGTGGAG ATGTGCCATG
24101 ACCTGTGTCC ATGAGCCCA CTGCACCCCT GGCAACAGG GCCCTCCCGT
24151 CCTTGGCTGG GCTGCGAGAT GGAGATGACA ACGGCCAAAG AACATTTGGG
24201 GAAGAACCAG CCATGCCACG AGCAGAGTCA GAAGTCCGAG GGGATAGAAT
24251 GCAGTCTCCC GTCCCCACC CACCCCTGTC CTCTGAATCA TGGCAGAAAC
24301 TAGCCTTCCA GCCCTCAGCA GCTCACATGG GGGACACGGC ACCCAAATCA
24351 CCACCAAGAA GTGCTGGCCG GTCTCTGCGA GGGCCACAGG CGCGGTGTA
24401 CTCGGTGGAA GTCTGGCGAT GTCAGAGACA GGCTCGGGGC AAGGACAGGT
24451 GTGGGGGTTT GAATAAGTGC ATTTGGGGAA CATGGCAGGG TGGTGACCTT
24501 TGTCTCTTCT CGAGACACTG GTGAGGTGTG GGTGCTGTCT GGTTCCTTGG
24551 ATCGCCCCC ACACTGGGGC AGAGTGGGAG ATGCTGGTGT GGGGGACATC
24601 AGCTCCACAC TCTGGGCCAG AGGGAGCCCC GGGGAAGGAA TGCTGAGGGC
24651 CCAGGGCCCT CGCTGGGAT CTGCACAGCT TTATAAGCAG CCCAGGGTGA
24701 GAGATGGGCC TGTCTGTGGT TCCCAGAGAC CACGGCAGGA AATTCTCTGT
24751 CACCATCGGT GCATGGGCAG GGGCCAGAGA ACCGGTGGCA CAAGGTGTCC
24801 CTGGCTCTCT GCTCAACAAA CAGCGAGTGC CCAGTGACTG CGACAGGGCC
24851 CCGCTCTTGG GATGAGGACA ACCGTCTGGG AACGTCCACG CACCTTTTAT
24901 GAGACACAGC ACGCGCCAGC ACCGCAGTCA CACACCGGGG GCTCGGGTCA
24951 GCCTCATAGC TGCCCGGCTT TGAGTGCTGG GCCTGCTCTT GTGAGCAGCG
25001 CCCACCTGGG TGGCGGTGGT GTGTGCTTCA CTTCCACACA CTCCCGTGCA
25051 TGCTCCCCGG CCTTCTGGGG TGGCTCGGGC TGTCGGGTGA GAATGTAGGC
25101 GGGGGGGGGG GGGGGCCTCT GTCCCGCCCT GGATGTTGGC TGCCCTCTGC
25151 CCGCCCTTGG ATGTTGGCTG CCCTGCCGGC CATCTTCCCT GTGAGAGGGT
25201 GCGCTCTTCC CTGCCATTGA GGGGAGAAGA GCTGCGGCTG CAGGAGTCTG
25251 CGGACCAAGC CAGGCCCGC CAGGCCCGCC CTGTGCAGAG ACGGCTGGG
25301 GGAGAGGAGA CGGGGCCCTT CTTCATGCA CAGGCGGCTT CAAACCCAGA
25351 CGTCTTTAAT GGGCCTGATT CACATCAGAG GCAGGATGAC TGCCTGTCCA
25401 GCGGTGGGT GGTTCCTGGC GGTTCCTGGC TACAGTGTCC TCAAGTGTACA
25451 AAGCTGTCTA TAAGAAGCCT ACGGAAATAC ACAATCTGTA ATAAGAGGAC
25501 AGTGTCTTCC TAAAGGATCG CAAAACCTTC CTGGATGAGG GCTACATGGA
25551 AGCTTAGGTG TGGGCCTTGG GGTGCGTAAA AGGGACCCCT CACGGGCGGG
25601 GCT

(SEQ ID NO: 3)
FEATURES:
Start: 3000
Exon: 3000-3118

FIGURE 3G

Intron: 3119-4126
Exon: 4127-4300
Intron: 4301-5310
Exon: 5311-5348
Intron: 5349-5478
Exon: 5479-5568
Intron: 5569-5758
Exon: 5759-5862
Intron: 5863-6571
Exon: 6572-6688
Intron: 6689-8108
Exon: 8109-8234
Intron: 8235-8322
Exon: 8323-8454
Intron: 8455-21992
Exon: 21993-22060
Intron: 22061-22942
Exon: 22943-23081
Stop: 23082

SNPs:

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
1931	G	A	Beyond ORF(5')			
4232	G	A	Exon	75	T	T
4309	G	T	Intron			
4622	G	A	Intron			
4708	T	C	Intron			
5787	C	T	Exon	150	N	N
5884	C	T	Intron			
8481	C	T	Intron			
8754	G	A	Intron			
8847	G	A	Intron			
11114	G	A	Intron			
19741	A	G	Intron			
20908	C	T	Intron			
22728	T	C	Intron			
23406	C	T	Beyond ORF(3')			
24078	G	A	Beyond ORF(3')			
24777	G	A	Beyond ORF(3')			

Context:

DNA
Position

1931 AGTTCAGTGGTACAGTCGTGGCTCACTGCAGCCTCAACCTCTCAGGCTCAAGCAATCCTC
CCTTCTCAGTCTCCTGAGTAGCTGGGACTACGGGTGTGTGCCACCACGCTTCCGGGCAT
GCAGCCAGGAGCCCAGGCCATCTGTGGCCACCTTGAGATCCAGAATCATCCATTTCT
CCAGGCCCCCTGCTGGGCTCCAACCTCTTGAGGACCAGAGAGCAGAGGTTGTGGAAGGCC
TTGGAAACGGGTCTGGATTACCTGTCTGGGAAGGTCTCTCCCAACCTGAGTGTGAGACA
[G,A]
GGGTTAGCTCTGCTGCTCACAATTTTGTGCCTTAATTCTGGCTTCCCTTTGGGGATCTT
CATCCTCAATTCTGATTGACATCCTTGCCACAAGGGACCCCCCTGCTCATTGATGCTTC
TCACCCGTCACCTCACTCTCATCCTCACTGCTAAGCAATTAGCCGTGTGTTTGCGGCATC
AGTGTTGACACCGATGATCCATGCTCAGAGGGTACAGGCCTGAAGAGCTATGTGGGGACT
GGCGCCCCGGAGGGGGTCCCGCTGTGGTGGCAGCGGTGGCCCCAAGCCCCACGCTCACT

4232 CCTCAGGATCTGGGCTCAGAGTGCTCTGGGCTGGGGTGCAATCAGTGCCTTGGGCAGGC
CCCTCCTCCTGGGAATGCCTGGTGGCTGATGCTGGGGTGGGGCTGTGGTCTTAGGGGGA
GTGTGTAGCTGTGGGAGCAGCCATGACTGGCTCCCCAGCTGTGCGCAACAGGCCTTC
CATCGGTGCCACAGGTACTCCTTCAGGACGAGGAGGACATGTTGATGGTCTGGACCT
GCTACTGGCGGGGACCTGCGCTACCACCTGCAGCAGAACGTGCAATTCTCCGAGGACAC
[G,A]
GTGAGGCTGTACATCTGCGAGATGGCACTGGCTCTGGACTACCTGCGCGGCCAGCACATC
ATCCACAGGTGTGTGCGTGGCAGACGGCGCAGGTACCTGCTGAGGTGGCGGGGCTGAAG
CAGCCTTAGGTGAGGCTGCCCGCACGGCGGCGTACTCCCTCAGAGCGGGTCTAGCTCCT
CTGCCCCACCTTTGCCTGAGTGCCTGCCCCAGCTGTGGCACCTGTGCGGACCAAGGTGAG
CCCCATAGCTGTGTGCTGGTGTCCATCTGGGGGGACCTGTCCTCCAGCAGCCCCAGCTGA

FIGURE 3H

4309 CCTGGTGGCTGATGCTGGGGTGGGGCTGTGGTCTTAGGGGGAGTGTGTCAGCTGTGGGA
GCAGCCATGACTGGCTCCCGAGCTGTGCGCACAAACAGGCCTTCCATCGGTGCCACAGGT
ACTCCTTCCAGGACGAGGAGGACATGTTTCATGGTCTGTGGACCTGCTACTGGGCGGGGACC
TGCGCTACCACTGTCAGCAGAACGTGCAGTTCTCCGAGGACACGGTGAGGCTGTACATCT
GCGAGATGGCACTGGCTCTGGACTACCTGCGCGGCAGCACATCATCCACAGGTGTGTGC
[G,T]
TGGCAGACGGCGCAGGTACCTGCTGAGGTGGGCGGGGCTGAAGCAGCCTTAGGTGAGGCT
GCCGGCACGGCGGCCGTACTCCCTCAGAGCGGGTCTAGCTCCTCTGCCCCACCCTTGCTT
GAGTGCCTGCCCCAGCTGTGGACCTGTGCCGACAGGTGAGCCCCATAGCTGTGTGCC
TGGTGTCCATCTGGGGGGACCTCGTCCAGCAGCCCCAGCTGAGACTGGGCACAGTGGGC
TGTTAGCCCTGGTGGACAGACCACAGGCTGGGTACAGCAGGTGGCTTACCTGGTCC

4622 CAGGTACCTGCTGAGGTGGGCGGGGCTGAAGCAGCCTTAGGTGAGGCTGCCGGCACGGCG
GCCGTACTCCCTCAGAGCGGGTCTAGCTCCTCTGCCCCACCCTTGCTTGAAGTGCCTGCC
CCAGCTGTGGCACCTGTGCCGACAGGTGAGCCCCATAGCTGTGTGCCGTGGTGTCCATCT
GGGGGGACCTCGTCCAGCAGCCCCAGCTGAGACTGGGCACAGTGGGCTGTTAGCCCTGG
TGGACAGACCACAGGCTGGGTACAGCAGGTGGCTTACCTGGTCCATTTAACTGAAG
[G,A]
CTCCTGTTTGCCCATCCACCACATCCAGGGAATCCAACTAATTTTAACTTAGCTTAA
AGCAGATGAAATTAGGAAGCAGAGCTGGTGTGATGGCTCTGAAATAAAATTTAAAAA
GAAATAGGAAGCAGATTATGAAGGAAGTGAATTTGGGAAGCAGAAATAGGCTGAAAT
CCGCAGCAATGGAAACAAATGAAATATCTGTGAGGTATATTTTAAAGTCGAATGGACTG
GTGTTTGCAATTTCTGCTTTGGGACTCGGATGTCTGATTATGACCTAGGCACCACTCAC

4708 CTCCTCTGCCCCACCCTTGCTTGAAGTGCCTGCCCCAGCTGTGGCACCTGTGCCGACCAG
GTACGCCCTTACCTGTGTCCCTGGTGTCCATCTGGGGGGACCTCGTCCAGCAGCCCCA
GCTGAGACTGGGCACAGTGGGCTGTTAGCCCTGGTGGACAGACCAGGCTGGGTGACA
GCAGGTGGCCTTACCTGGTCCATTTAACTGAAGACTCTGTTTGCCCATCCACCACATC
CCAGGGAATCCAACTAATTTTAACTTAGCTTAAAGCAGATGAAATAGGAAGCAGAGC
[T,C]
GGTGTGATGGCTCTGAAATAAAATTTAAAAAAGAAATAGGAAGCAGATTATGAAGGA
AGTGAATTTGGGAAGCAGAAATTAGGCTGAAATTCGCAGCAATGGAAACAAATGAAAT
ATCTGTGAGGTATATTTTAAAGTCGAATGGACTGGTGTGTTGCAATTTCTGCTTTGGGGAC
TCGGATGTCTGATTATGACCTAGGCACCACTGACTGAGCACTGGCTGTGTACCTGGAAAA
GTTGGGACAAAGCAAGAGCCGAGGTGGCTTGGTCTCTAGAGGCCGAGTCTTGAGGGGG

5787 CACCTGACCGACTTCAACATTGCCACCATCATCAAGGACGGGGAGCGGGCGACGGCATT
GCAGGCACCAAGCCGTACATGGGTGAGCCCCGAGCTGGGGTTCCAGATGGGAGCTGGCTTC
CTCCAGGTGGGAAGGACAAGACCTCGGTGGCTTCTCTGTCCACCCTGGAGGCGACCTGG
TCTCGGGATGTGGCCTCAAGGTGCCGGCCTGTGCCACGGGTCCGGGCTGTGACCCCTG
GGCAGCTGTTTTCTTCTTTCTGTGCGAAAGCTCCGGAGATCTTCCACTCTTTTGTCAA
[C,T]
GGCGGGACCGGCTACTCCTTCGAGGTGGACTGGTGGTGGTGGGGGTGATGGCCTATGAG
CTGCTGCGAGGATGGGTATGGACCCCTGCAGCCCCGGGCTTGGCTGCCAGGCCCTGC
TCTCTGCCCCACCACTGCTGGGAGGGGGTGGCTGCCCACTGCCCAGGTGCGCAGGGA
TGCTCTCACTGTGTCTGAGGAGTCACGCTTTTATCGAAGTGTGTAGTTGGTGTGGAATG
CCTGAGCAGGAGGAGGAAGGACAGACTCACTGTGGTTTCCCGGGCCGCTGCTGGTGCCT

5884 GGTTCAGATGGGAGCTGGCTTCTCCAGGTGGGAAGGACAAGACCTCGGTGGCTTCTCT
GTCCACCTCGGAGGACCTGGTCTCGGGATGTGGCCTCAAGGTGCCGCCCTGTGCC
ACGGGTCCGGGCTGTGACCCCTGGCAGCTGTTTTCTTCTTCTGTGCGAAAGCTCCG
GAGATCTTCACTCTTTTGTCAACGGCGGGACCGGCTACTCCTTCGAGGTGGACTGGTGG
TCGGTGGGGGTGATGGCCTATGAGCTGCTGCGAGGATGGGTATGGACCCCTGCAGCCCC
[C,T]
GGGCTTGGCTGCCAGGCCCTGCTCTGCCCCACCACTGCTGGGGAGGGGGTGGCTGC
CCCAGTGCCCAAGGTGCGCAGGGATGTCTCACTGTGTCTGAGGAGTCACGCTTTTATCGA
AGTGTGTAGTTGGTGTGGAATGCTGAGCAGGAGGAGGAAGGACAGACTCACTGTGGTT
TCCCGGGGCCGCTGCTGGTGCCTGCAGGCCAGCTCTGTGGGGGTGGCAAGGCTGAGAA
CTGGCCAGCAGGGGTGCTGCCTCGAACTTTCCACAAAAGTTTCTTTTGGGGCCTGTG

8481 GTGCTGTGGGACCACCTGAGCGAGAAGAGGGTGGAGCCGGGCTTCTGCCCCAACGTAAGC
CTGTGGGCGGCTCAGGTGGGGGGCCCTGGGGATGGATGGCGTCTCCACGGGCCGGGG
CTCAGACCCATCCTCTGTAGAAAGGCCGTCTGCACTGCGACCCACCTTTGAGCTGGA
GGAGATGATCCTGGAGTCCAGGCCCTGCACAAGAAGAAGCGCCTGGCCAAGAACAA
GTCCCGGGACAACAGCAGGGACAGCTCCAGTCCGTGAGTGCAGGGCAGGCTCAGGGCG
[C,T]
GGCGGGGGCTGGGCTTGGGGCTCCTCTTACCACCGAGCAAGGTGTGTGGGGACCCCTG
ACAGTGACACAGTCTCGGAAGTCCAGCAGACCGTTTCTGAACTCTGAGAAGGCCAGAG
ACCTCCCTTCTGCTTTTCCAGCCCCACCTCGCTCCTTATGAAGCAGGTGGGACAGGAC
AACCAGGGCTGGGGTTATGAGTGACCGGGGATGGCCATGTGAAGCCTTCTGTGCTTGC
GGTGTGCTGTTGGTGTGTGCGGGGACGGCTATGTGAAGCCTCACACTGCCCA

8754 CGTGAGTGCCAGGGCAGGCTCAGGGCGGGCGGGGGCTGGGCTTGGGGCTCCTCTCTAC
CACCGAGCAAGGTGTGTGGGGACCCCTGACAGTGACACAGTCTCGGAAGTCCAGCAGACC

FIGURE 3I

GTTTCCTGAAGTCCTGAGAAGGCCAGAGACCTCCCTTCTGCCTTTCCAGCCCCACCTC
GCTCCTTATGAAGCAGGTGGGCAGGGACAACCAAGGCTGGGGTTATGAGTGACGGGGAT
GGCCATGTGAAGCCTTCGTGCTTGCCAGGTGTGCTGGTGTGGTGTGTGCGGGGAC
[G, A]
GCTATGTGAAGCCCTCACACTCGCCAGGTGCGTCGGCATCAGGTATGTGTGCCGGGACA
GCCATGTGAAGCCCTCACACTCACCCAGGTGCGTCGGCATCAGTTGTGTGTGGGGACG
GCCATGTGAAGCCCTCACACTCGCCAGGTGTGCTGGCTTTGGTTGTGTGTGCGAGGGATG
GCCACATGAAGCCCTCACACTCGCCAGGTGCGTCAGCATCAGGTGTGTGTGCGGGGACG
GCCATGTGAAGCCCTCACACTCGCCAGGTGCGTTGATGTTGTGTGTGCGAGGGATGGCCA

8847 GCACACGTCTCGGAAGTCAGCAGACCGTTTCCTGAAGTCCTGAGAAGGCCAGAGACCTC
CCTTCTGCCTTTCCAGCCCCACCTCCTCCTTATGAAGCAGGTGGGCAGGGACAACCA
GGGCTGGGGTTATGAGTGACGGGGATGGCCATGTGAAGCCTTCGTGCTTGCCAGGTGT
GCTGGTGTGGTTGTGTGCGGGGACGGCTATGTGAAGCCCTCACACTCGCCAGGTGC
GTCGGCATCAGGTATGTGTGCCGGGACAGCCATGTGAAGCCCTCACACTCACCCAGGTGC
[G, A]
TCGGCATCAGTTGTGTGTGTGGGGACGGCCATGTGAAGCCCTCACACTCGCCAGGTGTG
CTGGCTTTGGTTGTGTGTGCGAGGGATGGCCACATGAAGCCCTCACACTCGCCAGGTGCG
TCAGCATCAGGTGTGTGTGCGGGGACGGCCATGTGAAGCCCTCACACTCGCCAGGTGCG
TTGATGTTGTGTGTGCGAGGGATGGCCATGTGAAGCCCTCACACTCACCCAGGTGCGTTGA
TGTGAGTTGTGTGTGCGAGGGACAGCCATGTGAAGCCCTCAGACTAGCCAGGTGTGTCGG

11114 AGACGCATCCCCCAGCCCCCTGCTTCGCTGCCGCTGGGGTCGGCCTGTGTGTGCGTCTC
CATCTCCTCCCCCTCTTCTCATAGGACACCAGGCATTGGATTTAGGGCCACCTGATCC
AGTATGGCCCCATCTTATCTTGATGATATCTGCAAAGACCTCACTTCAAATGAGGTAC
ATTCACAGGTACCCAGGATTAGAATTTGAGTGTGTCATTTTTGGGGACACAGTTTGGCC
ATACCACAGGATGTGGCTGATATTCACCAAGGAGTAGCTATGGTTGTGTGTGATGTCA
[G, A]
GGTGACGGTGATGACCTGGGTCCCTCGGTGGTCCCTTGCCCTGAGTCTGCCTGAGCC
TGTGGTGGATGTCCTGGGAAACTCTTGTGCTCAGCCCCGTGCGACCTCCTCAGACCTG
GTGGGCCCTGTGTTGCTCCTGGGCAGAGACGGGTGTGAGTCCCTCCTCACCATGATGT
GGGGGGCAGGGGTGGGGTCATGCCCTGGGTGCCCTGATTTTTGGGGGAAACAGGCCCCCC
AGTGGGTGAGGCTCCCATCTGCCCTCCTCCAGGACGGCTGCCGGCAGCCCTGGGTGT

19741 GTCAGAGGAAGAGCAGGGCAGATCCAGGCCAGGCACAGACCTGCCATGGGGTACTGCTG
GCGGCGCTACACACACCTTGGGAAGGGAGTCCCTGTGGAAGGGGTGCTGGTCACACATC
TAGGTGACACAGCCCCGGCTTGGGCGCTGCTCAGAGCCACCCCTCCAGATGGTTCTGGAG
CAGCTCCTCAGGCTTCTGGTGGCCTCTCTGCTAGGAAAACATGGCTGTGGACGTTGCGAG
GATGACCAACAGCCCCCTGCCACTGGGCTGCACACAGGGCCACGACGGGCGCTCATGTTCT
[A, G]
CATCACTGGCGCCACCCAGCCCCCTCCACCTTGTGTCGCTGTGAATCGCAGGATCCCA
GCGGCTCAGTCGGACCTCATTCTGAGTAGTCTGAGCCTTAGGTACCGTCACTTCTC
AGGCCGGCCCCAGTTTGCAGACTTGTCTGTCTATATCAGGGTTAGACCAGAGAGTGCTGA
GACACAGCAGATCACCCAGCCTGTCTCTTCTGATGACTAAGGACAGGTCCCTGCCAG
GATCGTACTCCTTTAGGGGAGGCCACAGTGACAGGGCAAAGCTGGAGGGAGAGAGCCA

20908 CCTGAGGTGAGGATCAAGACTGGCGTGGCCAACATGGTGAACCCCGTCTCTACAAAA
ATTAGCTGGGTATGGTGGTGGGCACCTGTAATCCAGCTACTTTGGGAGGCTGAGGCAGCA
GAATCACTTGAACCTGGGAGGCGGAGTTGGCAGTGAGCCAAGATCACGCCACTGCACTCT
AGTCTGGGTGACAGAGCGAGACCAAAAATAGGATCCATGGATAGCAGGCAAGAGTGTCCA
GGTGTTCGAGGCACAGACGACACTGTGACAGGGAAGAGTCCCCTTAGCCCTGGCTGGGCG
[C, T]
GTGAAAGCATGCTGTTGTCCGTCTCGGTGAACGCAGACTGTTGTACGCATTGCATAAC
GATGCTTCCGTCACTGGCCAATCGCATGGGGGGTGGTCCCGTAAGATGGTAACACTGGG
TTTTGCTGTACGTTTTGTATGTCTAGATAGGGTTGAGCGTTCTGGTGTGTACCACTCAC
ACGTCCCTCCGACCTTCAGAGCCAGCTCCCTCCCTCCAGGGCCTTGGCTGTGACGTG
GGTGACTTCCTATGGATCTGAGGTTCTGTGGTCTCACAAGTGGGCATCCTCTGGCCTCA

22728 CACAGATGACTGGCAGCAGGCAAGGCAGGCCAGAAATAGCAGCAGCTGCCACGGTGGGGC
CCAAGGGAGGATGGATGCTCCCTCTGCCCGCACGGGGCAAGGAGGGCTCCTGGAGGAGG
TGGGTCTGAGCTCTTATGGACAGGACGTGCGAGGCAGCAGTGCAGACGGCTGAGGGCAC
TGACTGGCACCCTGGGGATCAGACGACCGGGTGAAGAAATGAGGCTTAGCCGAGCCTCATT
CCCAAGTCACTGACCTATGGACCTGCACAGTCAAGCCTTTTCGGCTTCTGGCTGGAAACA
[T, C]
GCCGAGCCTCGCCAGCATGCTCACGTGCCCCACCCGTCGCCAGGCTCCCTGCCAGTGTG
TCGGGAGCATGGCCTCTCCAGCAGACACCGAGCCTGTGGCCACGTTTGGGCATCCACGC
CATGGCCTATCCCATGAGCCCGTGGGCAGGTGATGGGACCGTGAAGCCAGGGAGGTGGGG
GCATAACGCCCTCCATGTGTTCTTGCCACCCAGGCTGAAGAGGAGCCAGGACCTCCCGA
GGGAGCCTCTCCCGCCCTGAGTCCAGGGATGCTGCGGAGCCTGTGGAGGACGAGGCGG

23406 ACCCTTGAGTGCTTTGGAGACTCGGCTGCCAGAGGGAGGGCCATGGGCCGAGGCCTGG
CATTACGTTCCACCCAGCCTGGCTGGCGGTGCCACAGTGCCCCGGACACATTTACA
CCTCAGGCTCGTGGTGGTGCAGGGGACAAGAGGCTGTGGGTGACGGGGACACCTGTGGAG
GGCATTTCCTGGGCCCCGAGACCCGCTAGATGGAGGAAGCGCTGCTGGGCGCCCTC

FIGURE 3J

TTACCGCTCACGGGGAGCTGGGGCCATGGATGGGACAGGAGTCTTTGTCCCTGCTCAGCC
[C, T]
GGAGGGCTGTGCACGGCCCTCGTCACAAGGTGACCCCTTGCGACACAGGCCGCGGGTGCCCC
AGGCTCGGCTCAGTTCTTGGAGGTCAAGGGCATGGGTTGGGGTAGTGGGTGGGGAGGTGA
ATGTTTTCTAGAGATTCAAACCTGCTCCAGCAATTTCTGTATAGTTTTACCTCTGAGAAT
TACAATGTGAGAACCGCTCGATGTTGCATGTTCTGCGTACGTCTGTCTGCCTGGCCG
TCAGGCCGGTGCTGCCGTTTCTGGTTGGCCTGGACTTGGGGCAGCCAGTGGGGTGGGCA

24078 GGCTGTGGAAGGCAGAGACGGTCCTGGAGGCAGAGGAGCCAGGACAGCACCGTGCACCG
TGGAGCCGCGCAGTGCCGGGCGAGTGTGGCCCTCCATAAAGGGACGTATCCCTCTCAC
TGTGGCTGGGTGGTTCTGTGGTTGGAACGTGAACTAACTGGGTAAACGGCCTGTGTGCTT
CTCTCTGGTCTCGCTGGAGGAGGACGGGCTCAGCCCGTCAGCCAGCGCTCCAGACAGGC
CTGTGCTGGTTTCTCTGAGGAAATGGGTGTGGCGGGTCTGTGCCCTTCCAGGACAGC
[G, A]
GCCATAGTGGACATGTGCCTAGACCTGTGTCCATGAGCCCCACTGCACCCCTGGCAAACA
GGGCCCTCCCCTCCTTGGCTGGGCTGCGAGATGGAGATGACAACGGCCAAAGAACATTTG
GGGAAGAACC GGCCATGCCACGAGCAGAGTCAGAAAGTCCGAGGGGATAGAATGCAGCTTC
CCGTCCCCACCCACCCTGTCTCTGAATCATGGCAGAACTAGCCTTCCAGCCCTCAG
CAGCTCACATGGGGGACACGGCACCCAAATCACCACCAGGAAGGGTGGCCCGGTCTCTGC

24777 GGAACATGGCAGGGTGGTGCACCTTGTCTTCTCGAGACACTGGTGAGGTGTGGGTGCT
GTCTGGTTCCCTTGATCGCCCCCACACTGGGGCAGAGTGGGAGATGCTGGTGTGGGGGA
CATCAGCTCCACATCTGGGCCAGAGGGAGCCCCGGGAAGGAAATGCTGAGGGCCAGGG
CCTTCGCTGGGATCTGCACAGCTTTATAAGCAGCCCAGGGTGAGAGATGGGCCTGTCTG
TGGTCCCAGAGACCACGGCAGGAAATCTCTGTCAACATCGGTGCATGGGCAGGGGCCA
[G, A]
AGAACC GGTCACAAGGTGTCCCTGGCTCTCTGCTCAACAAACAGCGAGTGCCCAGTGA
CTGCGACCAGGCCCCGCTCTTGGGATGAGGACAACCGTCTGGGAACGTCCACGCACCCTT
TATGAGACACAGCACGCGCCAGCACCGCAGTCACACACCGGGGGCTCGGGTCAGCCTCAT
AGCTGCCCGGCCTTGAAGTGTGGGCTGCGTCTGTGAGCAGCGCCACCTGGGTGGCGGT
GGTGTGTGCTTCACTTCCACACACTCCCGTGCATGCTCCCCGGCCTTCTGGGGTGGCTCG

Chromosome mapping:
Chromosome 10